PTO/SB/21 (02-04)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	09/611,257
Filing Date	July 6, 2000
First Named Inventor	Terrance P. SNUTCH
Art Unit	1646
Examiner Name	N. S. Basi
Attorney Docket Number	381092000721

Total Number of Pages in This Submission	8

ENCLOSURES (*Check all that apply*)

- [X] Fee Transmittal Form (1 page + duplicate for fee processing)
- [ ] Fee Attached
- [ ] Amendment/Reply
- [ ] After Final
- [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [X] Information Disclosure Statement Supplemental (3 pages)
- [ ] Certified Copy of Priority Document(s)
- [ ] Response to Missing Parts/ Incomplete Application
- [ ] Response to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
- [ ] After Allowance communication to Technology Center (TC)
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [X] Other Enclosure(s) (please Identify below):

  Form PTO-1449 (1 page + duplicate)
  Copies of 7 References
  Return Receipt Postcard

Remarks

Customer No. 25225

SEP 27 2004

TECH CENTER 1600/2900

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm or Individual name	MORRISON & FOERSTER LLP Kate H. Murashige - 29,959
Signature	Kate H. Murashige
Date	September 20, 2004

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail, in an envelope addressed to: MS Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: September 20, 2004 Signature: Marian L. Christopher (Marian L. Christopher)

sd-220682

PTO/SB/17 (10-03)
Approved for use through 7/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

FEE TRANSMITTAL for FY 2004

Effective 10/01/2003. Patent fees are subject to annual revision.

[X] Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 180.00

Complete if Known	
Application Number	09/611,257
Filing Date	July 6, 2000
First Named Inventor	Terrance P. SNUTCH
Examiner Name	N. S. Basi
Art Unit	1646
Attorney Docket No.	381092000721

METHOD OF PAYMENT *(check all that apply)*

[ ] Check [ ] Credit Card [ ] Money Order [ ] Other [ ] None

[X] Deposit Account:

Deposit Account Number: 03-1952

Deposit Account Name: Morrison & Foerster LLP

The Director is authorized to: *(check all that apply)*

[X] Charge fee(s) indicated below [X] Credit any overpayments

[X] Charge any additional fee(s) or any underpayment of fee(s)

[ ] Charge fee(s) indicated below, **except for the filing fee** to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	770	2001	385	Utility filing fee	
1002	340	2002	170	Design filing fee	
1003	530	2003	265	Plant filing fee	
1004	770	2004	385	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	

SUBTOTAL (1) ($) 0.00

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

			Extra Claims		Fee from below	Fee Paid
Total Claims		-20** =		x		=
Independent Claims		-3** =		x		=
Multiple Dependent						=

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	86	2201	43	Independent claims in excess of 3
1203	290	2203	145	Multiple dependent claim, if not paid
1204	86	2204	43	** Reissue independent claims over original patent
1205	18	2205	9	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($) 0.00

***or number previously paid, if greater; For Reissues, see above*

FEE CALCULATION (continued)

3. ADDITIONAL FEES

SEP 27 2004

TECH CENTER 1600/2900

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge – late filing fee or oath	
1052	50	2052	25	Surcharge – late provisional filing fee or cover sheet.	
1053	130	1053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for *ex parte* reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	420	2252	210	Extension for reply within second month	
1253	950	2253	475	Extension for reply within third month	
1254	1,480	2254	740	Extension for reply within fourth month	
1255	2,010	2255	1,005	Extension for reply within fifth month	
1401	330	2401	165	Notice of Appeal	
1402	330	2402	165	Filing a brief in support of an appeal	
1403	290	2403	145	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive – unavoidable	
1453	1,330	2453	665	Petition to revive - unintentional	
1501	1,330	2501	665	Utility issue fee (or reissue)	
1502	480	2502	240	Design issue fee	
1503	640	2503	320	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	180.00
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	770	2809	385	Filing a submission after final rejection (37 CFR 1.129(a))	
1810	770	2810	385	For each additional invention to be examined (37CFR 1.129(b))	
1801	770	2801	385	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify)

*Reduced by Basic Filing Fee Paid

SUBTOTAL (3) ($) 180.00

SUBMITTED BY

(Complete *(if applicable)*)

Name (Print/Type)	Kate H. Murashige	*Registration No. (Attorney/Agent)*	29,959	*Telephone*	(858) 720-5112
Signature	Kate H. Murashige			*Date*	September 20, 2004

sd-220681

71 1646 #

PATENT
Docket No. 381092000721

CERTIFICATE OF MAILING BY "FIRST CLASS MAIL"

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: MS Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450, on September 20, 2004.

Marian Christopher

Marian Christopher

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the application of:

Terrance P. SNUTCH et al.

Serial No.: 09/611,257

Filing Date: July 6, 2000

For: MAMMALIAN T-TYPE CALCIUM CHANNELS

Examiner: N. S. Basi

Group Art Unit: 1646

SUPPLEMENTAL INFORMATION DISCLOSURE STATEMENT UNDER 37 C.F.R. § 1.97 & 1.98

MS Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Dear Sir:

Pursuant to 37 C.F.R. § 1.97 and § 1.98, Applicants submit for consideration in the above-identified application the documents listed on the attached Form PTO-1449. A copy of the documents is also submitted herewith. The Examiner is requested to make these documents of record.

sd-219629

Form PTO-1449

INFORMATION DISCLOSURE CITATION IN AN APPLICATION

(Use several sheets if necessary)

Docket Number 381092000721	Application Number 09/611,257
Applicant: Terrance P. SNUTCH et al.	
Filing Date July 6, 2000	Group Art Unit 1646
Mailing Date September 20, 2004	

U.S. PATENT DOCUMENTS

Examiner Initials	Ref. No.	Date	Document No.	Name	Class	Subclass	Filing Date If Appropriate
	1.		60/117,339				January 27, 1999
	2.		08/985,809				December 5, 1997
	3.	10/2001	6,309,858	Dietrich et al.	435	69.1	
	4.	03/2002	6,358,706	Dubin et al.	435	69.1	
	5.	03/2003	6,528,630	Williams et al.	536	23.1	
	6.	07/2003	2003/125269	Li	514	44	

FOREIGN PATENT DOCUMENTS

Examiner Initials	Ref. No.	Date	Document No.	Country	Class	Subclass	Translation YES	Translation NO
	7.	11/2000	00/70044	WIPO				

OTHER DOCUMENTS

(including author, title, Date, Pertinent Pages, Etc.)

Examiner Initials	Ref. No.	Title

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if citation considered, whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

Form PTO-1449	Docket Number 381092000721	Application Number 09/611,257
INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Applicant: Terrance P. SNUTCH et al.	
(Use several sheets if necessary)	Filing Date July 6, 2000	Group Art Unit 1646
	Mailing Date September 20, 2004	

COPY

U.S. PATENT DOCUMENTS

Examiner Initials	Ref. No.	Date	Document No.	Name	Class	Subclass	Filing Date If Appropriate
	1.		60/117,339				January 27, 1999
	2.		08/985,809				December 5, 1997
	3.	10/2001	6,309,858	Dietrich et al.	435	69.1	
	4.	03/2002	6,358,706	Dubin et al.	435	69.1	
	5.	03/2003	6,528,630	Williams et al.	536	23.1	
	6.	07/2003	2003/125269	Li	514	44	

FOREIGN PATENT DOCUMENTS

Examiner Initials	Ref. No.	Date	Document No.	Country	Class	Subclass	Translation YES	Translation NO
	7.	11/2000	00/70044	WIPO				

OTHER DOCUMENTS *(including author, title, Date, Pertinent Pages, Etc.)*

Examiner Initials	Ref. No.	Title

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if citation considered, whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

This Information Disclosure Statement is submitted:

☐ With the application; accordingly, no fee or separate requirements are required.

☐ Before the mailing of a first Office Action after the filing of a Request for Continued Examination under § 1.114. However, if applicable, a certification under 37 C.F.R. § 1.97(e)(1) has been provided.

☐ Within three months of the application filing date or before mailing of a first Office Action on the merits; accordingly, no fee or separate requirements are required. However, if applicable, a certification under 37 C.F.R. § 1.97(e)(1) has been provided.

☒ After receipt of a first Office Action on the merits but before mailing of a final Office Action or Notice of Allowance.

- ☐ A fee is required. A check in the amount of __ is enclosed.
- ☒ A fee is required. Accordingly, a Fee Transmittal form (PTO/SB/17) is attached to this submission in duplicate.
- ☐ A Certification under 37 C.F.R. § 1.97(e) is provided above; accordingly; no fee is believed to be due.

☐ After mailing of a final Office Action or Notice of Allowance, but before payment of the issue fee.

- ☐ A Certification under 37 C.F.R. § 1.97(e) is provided above and a check in the amount of __ is enclosed.
- ☐ A Certification under 37 C.F.R. § 1.97(e) is provided above and a Fee Transmittal form (PTO/SB/17 is attached to this submission in duplicate.)

Applicants would appreciate the Examiner initialing and returning the Form PTO-1449, indicating that the information has been considered and made of record herein.

The information contained in this Information Disclosure Statement under 37 C.F.R. § 1.97 and § 1.98 is not to be construed as a representation that: (i) a complete search has been made; (ii) additional information material to the examination of this application does not exist; (iii) the information, protocols, results and the like reported by third parties are accurate or enabling; or (iv) the above information constitutes prior art to the subject invention.

In the unlikely event that the transmittal form is separated from this document and the Patent Office determines that an extension and/or other relief (such as payment of a fee under 37 C.F.R. §1.17(p)) is required, Applicants petition for any required relief including extensions of time and authorize the Commissioner to charge the cost of such petitions and/or other fees due in connection with the filing of this document to **Deposit Account No. 03-1952** referencing 381092000721. However, the Commissioner is not authorized to charge the cost of the issue fee to the Deposit Account.

Dated: September 20, 2004

Respectfully submitted,

By: _____

Kate H. Murashige
Registration No. 29,959

Morrison & Foerster LLP
3811 Valley Centre Drive
Suite 500
San Diego, California 92130-2332
Telephone: (858) 720-5112
Facsimile: (858) 720-5125

E.J.

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

January 7, 1999

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***08/985,809***
FILING DATE: ***December 5, 1997***
PCT APPLICATION NUMBER: ***PCT/US98/23161***

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

T. Wallace

T. WALLACE
Certifying Officer

S P E C I F I C A T I O N

TO ALL WHOM IT MAY CONCERN:

Be it known that Edward Perez-Reyes and Leanne L. Cribbs, citizens of the United states of America, and resident at 320 South Birchwood Drive, Naperville, IL 60540-5033 and 1737 N. Natoma, Chicago, IL 60707, respectively, have invented a certain new and useful T-TYPE VOLTAGE-GATED CALCIUM CHANNELS AND METHOD OF USING SAME of which the following is a specification.

T-TYPE VOLTAGE-GATED CALCIUM CHANNELS AND METHOD OF USING SAME

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made with Government support under Grant Number HL58728 awarded by the National Heart, Lung, and Blood Institute of the National Institutes of Health. The United States Government may have certain rights in this invention.

TECHNICAL FIELD OF THE INVENTION

The present invention relates to molecular biology, and more particularly to cloned T-type calcium channels.

BACKGROUND OF THE INVENTION

Biological membranes are themselves generally impermeable to ionic species. Thus, ions enter cells through regulated pores formed from membrane-associated proteins. Most of these regulated pores are voltage-dependent and are thus able to transduce changes in the transmembrane potential into ion flux. Voltage-gated ion channels form a "superfamily" of related proteins (cf. Jan et al., *Nature, 345*, 672 (1990)). Peculiar to this genus is a high degree of conservation in molecular structure. Generally, voltage-gated channels are membrane bound glycolsylated proteins formed of many subunits. Large α subunits form a pore in the membrane that is selective for a given ionic species. Each α subunit contains four domains (I, II, III, and IV). Each channel domain has six putative transmembrane helical segments (S_1-S_6). In general, the segments within each domain are similar but not identical. Aside from overall structural conservation, certain charged residues within the domains are highly conserved among voltage-gated ion channels (Jan et al., *supra*; Stühmer et al., *Nature, 339*, 597-603 (1989)).

Differences in charged residues between groups of voltage-gated ion channels confer properties unique to each subgroup, such as ion selectivity. For example, most voltage gated ion channels are selective for either sodium, potassium or calcium. Known calcium channels require a ring of negative charge provided by glutamate residues found at similar locations in each of the domains (Yang et al., *Nature, 366*, 158-61 (1993)).

Voltage-gated channels are often classified on the basis of their electrophysiology. The resting membrane potential of most animal cells is between about –70 mV and –80 mV. When the membrane becomes depolarized (moved towards 0 mV), various membrane channels become activated (they are said to "open"). Thus, one category for classifying membrane channels is on the basis of the membrane potential necessary to

activate (or "gate") them (voltage dependency). For example, "T-type" calcium channels are activated at a lower voltage than L- or N-type channels (Nowycky et al., *Nature, 316*, 440-43 (1985)). Other physiological properties are the activation kinetics, inactivation kinetics, tail current (deactivation kinetics), and single channel conductance. Thus, in comparison to other calcium currents, T-type calcium current is characteristically short (Chen et al., *J. Gen. Physiol., 96*, 603-30 (1990)), and it exhibits characteristically slow activation kinetics near threshold, fast inactivation kinetics, and slow tail current (Randall et al., *Neuropharmacol., 63*, 879-93 (1997); Carbone et al., *Nature, 310*, 501-02 (1984); Nilius et al., *Nature, 316*, 443-46 (1985)).

Calcium currents have been implicated in many neurological and muscular functions. For example, T-type calcium current is associated with cardiac pacemaker activity, pain transmission in the central nervous system, and in other physiological functions. Defects in T-type calcium current have been implicated in cardiac arrhythmia, hypertension, and epilepsy. Given their potential clinical value, the pharmacological properties of calcium channels have been the subject of extensive study. Most such studies have involved L-type channels because, unlike T-type channels, L-type calcium channels are readily purified from cell extracts. For example, L-type calcium channels have been purified using dihydropyridine drugs (e.g., nifedipine) which can bind with sufficiently high affinity to serve as a ligand for purifying L-type calcium channels. Such purified and cloned L-type calcium channels have been used to develop assays for drugs affecting L-type calcium channels (see, e.g., U.S. Patents 5,429,921 and 5,386,025).

While many electrophysiological characteristics of T-type calcium currents are known, the lack of isolated T-type channels has stalled research into the pharmacology and biophysics underlying the T-type calcium current, at least in comparison with other calcium channels. Indeed, while it is generally assumed that voltage-sensitive ion channels are responsible for the current, no such channel protein, nor any nucleic acid encoding such a protein, has been isolated. In view of the foregoing problems, there exists a need for an isolated T-type calcium channel and a nucleic acid encoding a T-type calcium channel.

BRIEF SUMMARY OF THE INVENTION

The present invention provides an isolated or substantially purified nucleic acid encoding a protein comprising at least one domain of a T-type calcium channel and cells expressing such nucleic acids. The present invention also provides an isolated or substantially purified T-type calcium channel and an isolated or substantially purified antibody molecule recognizing an epitope on a T-type calcium channel protein.

The present invention is useful for exploring the electrophysiology and pharmacology of the T-type calcium current. Such knowledge can lead to the development of drugs for potentiating or attenuating T-type calcium channels. Thus, the present invention provides an assay for identifying potential drugs affecting T-type calcium channels by exposing cells expressing a T-type calcium channel to a putative drug and then measuring the calcium flux in response to a change in membrane potential. The identification of drugs affecting T-type calcium channels will facilitate even greater understanding of the biophysics of these proteins. Furthermore, some such drugs could have potential clinical applications.

The invention can best be understood with reference to the accompanying drawings and in the following detailed description of the preferred embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1G show the complete nucleotide and amino acid sequences (SEQ ID NO:1 and SEQ ID NO:2) of a T-type calcium channel (α1G or C_{av}T.1)), and the conserved functional domains.

Figures 2A-2F show the complete nucleotide and amino acid sequences (SEQ ID NO:3 and SEQ ID NO:4) of a T-type calcium channel (α1H or C_{av}T.2), indicating conserved functional domains.

Figure 3 compares the amino acid sequences of domains of the T-type calcium channels with those of other calcium channels.

Figures 4A-4D are graphic representations of the current-voltage relationships of two cloned T-type calcium channels (Figures 4A and 4B), a native T-type calcium current in NIE-115 cells (Figure 4C), and a cloned R-type calcium channel (Figure 4D).

Figure 5A is a graphic representation of the average current-voltage curve for cloned T-type calcium channels (α1G, closed circle, α1H, open circle), a native T-type calcium current in NIE-115 cells (triangles), and a cloned R-type calcium channel (filled squares). Figures 5B and 5C are graphic representations of the conductance of calcium channels. Figure 5B compares the conductance in 2 mM $BaCl_2$ of cloned T-type calcium channels (α1G, closed circle, α1H, open circle), a native T-type calcium current in NIE-115 cells (triangles), and a cloned R-type calcium channel (filled squares). Figure 5C compares the normalized conductance of a cloned T-type calcium channel at three different concentrations of $BaCl_2$.

Figures 6A and 6B are graphic depictions of the kinetics of a cloned T-type calcium channel. Figure 6A compares the current recorded in cells expressing cloned T-type (α1G) or L-type (α1E) calcium channels at -20 mV. Figure 6B compares the

voltage dependent time constants of cloned T-type calcium channel activation and inactivation.

Figures 7A-7F are graphic depictions of the tail current of a cloned T-type calcium channel. Figures 7A and 7D depict tail current amplitudes for α1G and α1H, respectively. Figures 7B and 7E depict tail current at several test potentials for α1G and α1H, respectively. Figures 7C and 7F depict average kinetics of the tail current as a function of repolarization potential for α1G and α1H, respectively.

Figures 8A-C graphically illustrate the voltage dependence of the inactivation of a cloned T-type calcium channel. Figure 8A illustrates the inactivation of cloned T-type calcium channels due to 200 ms pre-pulses at various potentials followed by a test pulse to -20 mV. Figure 8B compares the inactivation of cloned T-type (circles) and R-type (squares) calcium channels due to 200 ms pre-pulses at various potentials followed by a test pulse to -20 mV in comparison to a -100 mV control. Figure 8C depicts the voltage dependence of inactivation induced by 10 s pre-pulses for cloned T-type (circles) and R-type (squares) calcium channels.

Figures 9A-9C graphically illustrate the single channel conductance of a cloned T-type calcium channel. Figure 9A depicts the raw data collected from a patch of membrane on an oocyte expressing a cloned T-type calcium channel at various voltage protocols. Figure 9B represents the ensemble current recorded from 100 sweeps. Figure 9C graphically illustrates the single channel amplitude plotted against test potential.

Figures 10A and 10B graphically present data concerning the use of a cloned T-type calcium channel to detect drugs affecting the channel. Figure 10A depicts the effect of 100 μM on current-voltage relationships with a single dosage of miberfradil. Figure 10B illustrates the effect on T-type channel conductance of various doses of miberfradil.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention provides an isolated or substantially purified nucleic acid encoding a protein comprising at least one domain of a T-type calcium channel α subunit. The nucleic acid can be of any type, and it can include other elements aside from a sequence encoding a T-type calcium channel domain or domains. For example, where the nucleic acid comprises RNA, it can also include regulatory sequences suitable to permit translation of the RNA. Thus, an RNA nucleic acid of the present invention preferably has at least one ribosome entry site, and preferably has a poly-adenosine tail for stabilizing the RNA in the cellular environment.

Similarly, DNA nucleic acids of the present invention can have regulatory elements for promoting the transcription of sequence encoding the T-type calcium

channel into an RNA such as that described above. For example, a DNA nucleic acid of the present invention can have a promoter and/or an enhancer sequence.

The choice of promoter and/or an enhancer will largely depend on the milieu in which the nucleic acid is to be expressed. Thus, for expression in bacterial cells, the regulatory elements are bacterial promoters. Similarly, for expression in mammalian cells, the regulatory elements are able to effect expression in mammalian cells.

While many such regulatory elements are known in the art, examples include prokaryotic promoters and viral promoters (e.g., retroviral ITRs, LTRs, immediate early viral promoters (IEp), such as herpesvirus IEp (e.g., ICP4-IEp and ICP0-IEp), cytomegalovirus (CMV) IEp, and other viral promoters, such as Rous Sarcoma Virus (RSV) promoters, and Murine Leukemia Virus (MLV) promoters). Other suitable promoters are eukaryotic promoters, such as enhancers (e.g., the rabbit β-globin regulatory elements), constitutively active promoters (e.g., the β-actin promoter, etc.), signal specific promoters (e.g., inducible promoters such as a promoter responsive to RU486, etc.), and tissue-specific promoters (e.g., those active in epidermal tissue, dermal tissue, tissue of the digestive organs (e.g., cells of the esophagus, stomach, intestines, colon, etc., or their related glands), smooth muscles, such as vascular smooth muscles, cardiac muscles, skeletal muscles, lung tissue, hepatocytes, lymphocytes, endothelial cells, sclerocytes, kidney cells, glandular cells (e.g., those in the thymus, ovaries, testicles, pancreas, adrenals, pituitary, etc.), tumor cells, cells in connective tissue, cells in the central nervous system (e.g., neurons, neuralgia, etc.), cells in the peripheral nervous system, and other cells of interest). While the nucleic acid can be any type of nucleic acid, the nucleic acid preferably comprises a cDNA. A cDNA nucleic acid is preferred over other nucleic acids to permit the nucleic acid to be readily cloned, sequenced, and expressed in a wide variety of cells.

The isolated or substantially purified nucleic acid of the present invention encodes all or part of a T-type calcium channel α subunit. As used herein, a "calcium channel" includes a protein structure for facilitating the flux of calcium ions across a biological membrane into which the calcium channel is inserted. As used herein, a "T-type channel" is a type of voltage-gated ion channel that facilitates the flux of ions when the membrane potential of a biological membrane into which it is inserted experiences a slight depolarization. Thus, a T-type calcium channel can gate at about -45 mV to about -30 mV (i.e., about -40 mV to about -35 mV) in 2 mM Ba^{2+}. Additionally, T-type channels of the present invention exhibit a slow deactivation (tail current) following depolarization. Thus, a T-type calcium channel can exhibit a tail current that decays exponentially with a tau value from about 2 ms to about 10 ms (e.g., from about 4 ms to about 7 ms, such as about 6 ms) following repolarization to a membrane potential from about –80 mV to

about –60 mV in a solution with a barium concentration of from about 10 mM to about 40 mM. Another defining characteristic of T-type calcium channels is that they exhibit small single channel conductance. Thus, for example, a T-type channel exhibits a single channel conductance of from about 7 pS to about 10 pS (e.g., from about 7.5 pS to about 9.5 pS), and typically from about 8 pS to about 9 pS in a solution with a barium ion concentration of about 0.1 M.

The isolated or substantially purified nucleic acid of the present invention encodes all or part of any T-type calcium channel having at least one of the aforementioned electrophysiological properties when properly assembled within a cellular membrane. The general structure of calcium channels is summarized above and is otherwise known in the art. Thus, for example, the nucleic acid can encode one of the four functional domains mentioned above. As used herein, a domain of a T-type calcium channel is any protein structure able to associate with three other domains to form a tetrameric body functioning as a T-type calcium channel. While the native T-type calcium channel structure includes all four domains in a single polypeptide (indicated in Figures 1A-G and 2A-2F), a domain can exist as a polypeptide species separate from those containing the other domains. Such separate domains are able to associate within the plasma membrane to form a functional channel. Alternatively, where a plurality of domains are linked within a common polypeptide, the linkage can deviate substantially from the native linkage. Thus, for example, the domains can be linked by polypeptide sequences other than those sequences (see, e.g., Figures 1A-1G and 2A-2F) linking the domains in the native protein (e.g., non-native polyglutamate linkages). Indeed, the domains themselves can include non-native linkages between membrane-spanning elements within the domains. Aside from these modifications, the nucleic acid can encode a chimeric calcium channel domain (or an entire channel) comprising a portion of a T-type calcium channel and a portion derived from another calcium channel (or other channel) protein. For example, the chimera can include portions of domains from T-type channels responsible for low voltage gating and portions of domains from other calcium channels responsible for slow inactivation. Such a protein exhibiting T-type gating but longer inactivation kinetics would facilitate pharmacological research.

As mentioned, nucleic acids of the present invention can encode an entire T-type channel (i.e., a T-type channel protein comprising four functional domains). Examples of the amino acid sequences of two full-length T-type channels are set forth at SEQ ID NO:1, SEQ ID NO:3, and examples of sequences encoding full length T-type calcium channels are SEQ ID NO:2 and SEQ ID NO:4. However, the invention is not limited to these exemplary sequences. Indeed, as mentioned, an amino acid sequence of a T-type calcium channel can vary from those listed, and it is within the state of the art to change a

nucleotide sequence encoding a T-type channel to introduce mutations into the protein. For example, mutations comprising insertions or deletions can be introduced on either the amino- or carboxy-terminus of the protein, or such mutations can be intrasequence insertions or deletions. Where the electrophysiological properties of the calcium channel are to be conserved, such mutations preferably are in regions other than the membrane spanning domains (see, e.g., Figures 1A-1G And 2A-2F). For example, SEQ ID NO:5 and SEQ ID NO: 6 are the sequences of two T-type channels having deletions in the region linking domains III and IV. However, in some applications (e.g., to decrease inactivation kinetics), the changes can be within the membrane-spanning regions. Moreover, as mentioned above, the sequence can form a protein having only one functional domain of a T-type calcium channel. Additionally, the sequence can also form a chimeric protein or domain, such as those described above.

Aside from insertions and deletion mutations of native T-type calcium channel sequences, a T-type calcium channel can include substitutions of amino acid residues, e.g., for those indicated in SEQ ID NO:1 and SEQ ID NO:3. Preferably, and especially where such a substitution is within a membrane spanning region, the substitution is conservative. Thus, within membrane spanning domains, positively-charged residues (H, K, and R) preferably are only substituted with positively-charged residues; negatively-charged residues (D and E) preferably are only substituted with negatively-charged residues; neutral polar residues (C, G, N, Q, S, T, and Y) preferably are only substituted with neutral polar residues; and neutral non-polar residues (A, F, I, L, M, P, V, and W) preferably are only substituted with neutral non-polar residues. Figure 3 indicates the conservation between the S-IV domains of T-type calcium channel α subunits and those of other calcium channels. Preferably, any amino-acid substitution within the membrane-spanning regions does not alter this conservation. Most preferably, any substitution, deletion, or insertion does not alter the IVs4 domain. In each of the exemplary T-type calcium channel α subunit sequences (SEQ ID NO:1 and SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:6), the putative S4 region comprises Arg Ile Met Arg Val Leu Arg Ile Ala Arg Val Leu Lys Leu Leu Lys Met Ala Val Gly Met Arg Ala (SEQ ID NO:7). Given the strong sequence conservation among families of voltage-gated ion channels, it is likely that SEQ ID NO:7, or a derivative sequence, will be present in T-type channels. Thus, the present invention provides any T-type calcium channel (or a nucleic acid encoding such a T-type calcium channel) comprising SEQ ID NO:7 or a sequence derived from SEQ ID NO:7 having conservative amino acid substitutions, as described above.

The nucleic acid of the present invention encoding all or a part of a T-type calcium channel can be isolated via any suitable method. For example, prior to the present

invention, one of skill in the art could design a probe based on the sequence of known, non-T-type, calcium channels and use such probe to screen a genetic library. If such a screen were to identify a putative calcium channel, the researcher could then attempt to clone the entire nucleic acid to characterize it. Similarly, prior to the present invention, to isolate a nucleic acid encoding a T-type calcium channel, one of skill in the art could consult publicly available databases containing DNA sequences (e.g., Genbank) to locate nucleic or amino acid sequences representing a portion of a T-type calcium channel protein or nucleic acid. However, such databases contain no sequence for a full-length T-type calcium channel. Such methods assume that T-type calcium channels share sufficient sequence identity with known calcium channel nucleic acids to cross-hybridize, an assumption not supported by any published report. Moreover, prior to the present invention, no partial sequence in such databases was identified as corresponding to a T-type calcium channel. Thus, prior to the present invention, the presence of partial sequences in the public DNA databases could facilitate the isolation of T-type calcium channels only with the exercise of a considerable degree of speculation on the part of the researcher.

By providing several sequences pertaining to T-type calcium channels and a comparison presenting conserved regions and domains, the present invention greatly facilitates the isolation of other nucleic acids encoding T-type calcium channels (or derivatives thereof) with much less experimentation. Thus, while any of the methods discussed above can be employed to isolate other members of this genus, preferably, a nucleic acid encoding a T-type calcium channel is isolated by probing a genetic library using a probe that hybridizes to a DNA encoding a peptide sequence contained in (or similar to) a known T-type calcium channel (e.g., SEQ ID NO:1 or SEQ ID NO:3). To facilitate the isolation of a T-type calcium channel, the present invention provides an isolated polynucleotide hybridizing to a portion of the nucleic acid of the present invention encoding a T-type calcium channel (or a portion thereof). Thus, for example, the present invention includes an isolated polynucleotide hybridizing to SEQ ID NO:2 or SEQ ID NO:4. The isolated polynucleotide can hybridize to all or any portion of the sequence encoding the T-type calcium channel.

To isolate such a polynucleotide, any portion of a sequence encoding a T-type calcium channel can be employed as a probe to screen a genetic library, and such screening can be accomplished by standard techniques known in the art. While the probe can hybridize to any portion of such a DNA, preferably the probe is designed to hybridize to a DNA encoding a polypeptide sequence that is highly conserved among T-type calcium channels but is less conserved between the genus of T-type calcium channels and other proteins. Such peptide sequences are readily apparent from the sequence

comparison set forth in Figure 3. Generally, the specificity of hybridization in a genetic screen varies depending on the length of the probe and the stringency (e.g., temperature, salt and detergent concentration, etc.) of hybridization. Stringency of hybridization is broadly classified as "high," "moderate," or "low," and the parameters of these terms are well recognized in the art (see, e.g., Sambrook et al., "Molecular Cloning, a Laboratory Manual," Cold Spring Harbor Press, 1989). The isolated polynucleotide hybridizing to a portion of the nucleic acid encoding a T-type calcium channel can hybridize under any desired stringency conditions. However, for identifying other T-type channels, preferably, the hybridization occurs under moderate stringency, and most preferably under high stringency.

Of course, the isolated or substantially purified polynucleotide can itself be employed as a probe to screen a library as described to isolate a second nucleic acid. In such a screen, one of the polynucleotides will be complementary to a portion of the sequence encoding the T-type calcium channel, and the other isolated nucleic acid will be "sense." Preferably, one of the two isolated polynucleotides (the "sense" strand) itself encodes a T-type calcium channel, or at least one domain thereof. Such a sequence can be cloned to be operably linked to suitable regulatory elements, as described, to produce a T-type calcium channel. Thus, aside from using the nucleic acid of the present invention to produce a T-type calcium channel, the nucleic acids of the present invention are also useful for isolating other sequences encoding T-type calcium channels, or derivatives thereof.

However isolated, the isolated or substantially purified nucleic acid of the present invention is useful, in part, for producing all or a portion of a T-type calcium channel. Thus, the nucleic acid can be introduced into a suitable milieu for driving its expression. Because T-type channels are transmembrane proteins, preferably such a milieu is a living cell. However, it should be understood that the nucleic acid can also be expressed *in vitro* under conditions, such as those known in the art, suitable for *in vitro* transcription and translation. However produced, the present invention includes any protein, such as a recombinant protein or an isolated or substantially purified protein, including all or a portion of a T-type calcium channel or a protein derived from a T-type calcium channel. Such proteins are described above.

For expression in a living cell, the nucleic acid must be introduced into the cell. As nucleic acids are generally introduced into cells as part of genetic vectors, the present invention provides a vector having a T-type calcium channel nucleic acid of the type described above. Any type of vector suitable for introducing the nucleic acid into a host cell is within the context of the present invention. Examples of such vectors include naked DNA and RNA vectors (such as oligonucleotides, plasmids, capped cRNA, etc.),

viral vectors such as adeno-associated viral vectors (Berns et al., *Annals of the New York Academy of Sciences, 772*, 95-104 (1995)), adenoviral vectors (Bain et al., *Gene Therapy, 1*, S68 (1994)), herpesvirus vectors (Fink et al., *Ann. Rev. Neurosci., 19*, 265-87 (1996)), packaged amplicons (Federoff et al., *Proc. Nat. Acad. Sci. USA*, 89, 1636-40 (1992)), pappiloma virus vectors, picornavirus vectors, polyoma virus vectors, retroviral vectors, SV40 viral vectors, vaccinia virus vectors, and other vectors. Once a given type of vector is selected, its genome must be manipulated for use as a background vector, after which it must be engineered to incorporate exogenous polynucleotides. Such manipulations are known in the art.

The vectors of the present invention are useful for introducing a nucleic acid encoding all or a portion of a T-type calcium channel into a host cell. Thus, the present invention provides a cell into which the vector of the present invention has been introduced. The host cell can be any cell suitable for expressing the nucleic acid (e.g., bacteria, insect cells, mammalian cells, etc.). The host cell can thus be *in vitro* or *in vivo*. Preferably the cells do not exhibit native T-type calcium current. A preferred cell type is HEK-293 cells because they contain genetic elements that facilitate the expression of transgenes from a variety of expression vectors. For facilitating electrophysiological recordings, oocytes (e.g., *Xenopus* oocytes) are preferred, as they are large and readily handled.

The vector can be introduced into the cell in any manner suitable for the cell type and vector employed. In one embodiment, the vector can be used to prepare an RNA transcript *in vitro* (e.g., a capped cRNA) which is then introduced into the host cell by standard methods (such as injection). Such techniques are preferred when the host cells do not actively transcribe DNA (such as oocytes). In other embodiments, a DNA vector is introduced into the cell such that it is transcribed within the cell. For example, the vector can be introduced into the cell such that it forms an extrachromosomal segment of genetic material in the cell, as is the case with many types of viral vectors. Alternatively, the vector can introduce the nucleic acid into the chromosomal DNA of the host cell. In this respect, a cell line comprising chromosomes into which the T-type calcium channel nucleic acid has been introduced is able to propagate the nucleic acid through several passages (e.g., for at least 10 passages), and, preferably, the nucleic acid is stably integrated into the chromosomes of such cells. Thus, the cell line can propagate the nucleic acid for at least 20 passages, and more preferably significantly more than 20 passages (e.g., at least about 25 passages, or even more).

Preferably, a cell into which the nucleic acid is introduced is also able to express the nucleic acid. The expression of the nucleic acid can be detected by probing the cell for the presence of T-type calcium channel mRNA, such as via Northern hybridization

analysis, in situ hybridization, etc. More preferably, however, the cell is able to express the nucleic acid to produce the protein including all or a portion of a T-type calcium channel. In such cells, expression of the nucleic acid is confirmed by detecting the protein, for example, by probing cellular extracts with an antibody recognizing the protein (e.g., on a Western blot, etc.). Of course, the protein contributes to the formation of a functional calcium channel in the membrane of the cell producing the protein. Where the protein encodes an entire α subunit, the full protein will possess some or all of the electrophysiological properties of T-type calcium channels described above. Where the protein encodes less than an entire channel α subunit (e.g., a domain), the protein will aggregate with other constituent domains in the membrane to form a functional channel. Thus, the presence of the protein can be detected by assaying the cell for T-type calcium channel activity. Indeed, assaying for channel activity serves to determine whether a nucleic acid encoding a putative calcium channel, in fact, encodes a species of T-type channel (as opposed to a member of another genus of calcium channels). For example, when large cells (e.g., oocytes) are used as the host cells, the electrophysiological properties of the channel can be investigated. Thus, the membrane activity of whole cells expressing the nucleic acid can be measured directly, such as via patch clamp techniques using a voltage clamp electrode and a current electrode (Bernal et al., *J. Pharmacol. Exp. Ther., 282*, 172-80 (1997)). Alternatively, the activity of single channels can be measured, such as with a standard depolarizing bath and pipette solutions (Lacerda et al., *Biophys. J., 66*, 183-43 (1994)). However measured, the properties of cells into which the putative nucleic acid is introduced are compared to the known channel conductance, voltage dependency, activation kinetics, inactivation kinetics, and tail current known for T-type channels and discussed above.

Regardless of the cell system, the ability to express a T-type calcium channel nucleic acid within host cells to produce an active channel permits the channel to be further studied. In this regard, the present invention provides a method of identifying a drug which affects T-type calcium channels. The method involves first expressing a T-type calcium channel in a cell to produce an active channel, as herein described. The cell expressing the channel is then exposed to a solution containing a putative drug for interfering with the channel. Thereafter, the presence or absence of calcium flux in response to a change in membrane potential is assayed. A quick method of assaying for calcium flux is first to introduce a calcium-sensitive labile dye into the cells. For example, the dye can be one such as those that fluoresce or change color in the presence of calcium (e.g., Indo-1). Thereafter, the cells are exposed to a depolarizing solution containing high (e.g., about 50 mM) potassium concentration and a drug, and the reaction of the labile die is compared to control cells. Using a labile dye affords the ability to

assay many putative drugs quickly in a high throughput assay for putative drugs affecting T-type channels. For example, the initial screening can be carried out in 96 well plates. Moreover, dose-response data can be readily generated by exposing the cells to several concentrations of the same putative drug.

Once a putative drug is detected, its effect on the electrophysiology of the cell (e.g., single channel conductance, voltage dependency, activation kinetics, inactivation kinetics, and tail current of the cells) can be investigated in detail. Generally, the effect of the putative drug on T-type calcium currents is assessed by measuring the various electrophysiological parameters in the presence of various concentrations of the drugs and comparing the data to untreated (or sham-treated) control cells. Cells preferably are maintained in a continuous perfusion chamber during such experiments to facilitate changing solutions. The inventive method of identifying a drug which affects T-type calcium channels can employ any nucleic acid encoding a T-type calcium channel (or derivative thereof), such as those nucleic acids described herein. In fact, as several isoforms of T-type channel exist (e.g., α1G and α1H), the assay method can be repeated using nucleic acids encoding different isoforms to identify drugs that preferentially target a given isoform, or drugs which affect more than one isoform of T-type calcium channels.

Aside from affording an *in vitro* assay for detecting potential therapeutic or investigative drugs targeting T-type calcium channels, the method of expressing the T-type calcium channel nucleic acid can also be used *in vivo*. For example, as mentioned, several neurological and muscular diseases or disorders have implicated mutations affecting native nucleic acids encoding T-type calcium channels. The present invention, thus, provides a method of treating a disease or disorder associated with a deficiency in a native T-type calcium channel nucleic acid. The method involves introducing a vector having the T-type calcium channel nucleic acid into cells of a host in which native expression of the nucleic acid is deficient. Thus, for example, for treating cardiomyopathy associated with deficiencies in T-type calcium channels, the vector is introduced into myocardial cells. Similarly, for treating forms of epilepsy associated with deficiencies in T-type calcium channels, the vector is introduced into neurons (e.g., thalamic neurons). Within the target cells, the nucleic acid within the vector is expressed to produce active T-type calcium channel. By similar methods, an nucleic acid having a sequence antisense to a sequence encoding a T-type calcium channel (or a portion thereof) can be expressed within a cell. The presence of an antisense sequence can down-regulate the expression of native T-type calcium channel genes by hybridizing to T-type channel mRNA within the cell. Thus, the present invention is useful to treating disorders associated with over-expression of T-type calcium channels.

T-type channel proteins (such as whole T-type calcium channels, domains of such channels, chimeras including portions of T-type calcium channels, etc.) can be employed to generate antibodies (e.g., immunoglobulins) to T-type calcium channels. Thus, the present invention provides an isolated and substantially purified antibody molecule recognizing an epitope on a T-type calcium channel. Such antibodies can be monoclonal antibodies or polyclonal antisera. Such antibodies can be produced by any suitable method, many of which are well known in the art. Antibodies recognizing T-type calcium channels can be used to purify the channels from cell extracts or other solutions by standard methodologies (e.g., immunoprecipitation). Moreover, depending on the location of the epitopes for the antibodies on the T-type calcium channel, the antibodies can be used to affect the channel proteins present on the surface of cells. Thus, antibodies directed to T-type calcium channels are potential reagents for studying the channels as well as for therapy.

EXAMPLES

Several examples are presented below to illustrate the invention. Taken together, the examples demonstrate the cloning of two novel proteins and their characterization as T-type calcium channel α subunits. These examples are included here for purely illustrative purposes; as such, they are not to be construed so as to limit the scope of any aspect of the invention.

Many procedures employed in the following examples are techniques routinely performed by one of ordinary skill in the art (see generally Sambrook et al., *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989)) and are not discussed in detail. However, some reagents and methods deserve specific description. Thus, for example, i*n vitro* translation and expression were conducted as described previously (Schneider et al., *Receptors and Channels, 2*, 255-70 (1995)). *Xenopus laevis* oocytes were prepared as described previously (Bernal et al., *J. Pharmacol. Exp. Ther., 282*, 172-80 (1997)). To express proteins, 10 or 30 ng of capped cRNA was injected into the oocytes or NIE-115 cells in a volume of 50 nl. For single channel recording, oocytes were injected with 100 ng capped cRNA and incubated for one week prior to assay.

Cells were voltage clamped using a two-microelectrode voltage clamp amplifier as described (Bernal et al., *J. Pharmacol. Exp. Ther., 282*, 172-80 (1997)). The standard bath solution contained the following: 40 mM $Ba(OH)_2$, 50 mM NaOH, 1 mM KOH, 0.1 mM EDTA, and 5 mM HEPES, adjusted to pH 7.4 with methanesulfonate. The osmolality of the 2 and 10 mM Ba^{2+} solutions was balanced by increasing the NaOH concentration as described (Lory et al., *J. Physiol., (London), 429*, 95-112 (1990)).

Voltage and current electrodes (1.5-1.8 M tip resistance) were filled with 3 M KCl. Except as noted, data were acquired at 4 kHz using the pCLAMP system, and filtered at 1 kHz. Data were analyzed using pCLAMP software. Boltzman fits and linear regression were calculated using Prism.

EXAMPLE 1

This example demonstrates the cloning and characterization of two putative T-type calcium channels.

A search of the Genbank library was conducted to identify clones identified as having some degree of homology to calcium channels. The search identified an expressed sequence tagged (EST) partial sequence in a human brain clone (H06096), which was used as a probe to screen a λgt10 cDNA library prepared from rat brain. Successive screening of the cDNA library identified five overlapping clones which were aligned to construct an entire cDNA sequence, termed α1G (representing nucleotides 379-7540 of SEQ ID NO:2).

The α1G cDNA was cloned into the pSP72™ vector and sequenced by standard computer-assisted sequencing. Using the α1G cDNA, the amino acid sequence of the α1G protein was deduced (SEQ ID NO:1) and compared to the sequences of other known calcium channel α subunits. Figure 1 sets forth these sequences and subunits, and it indicates the putative transmembrane domains of the protein. By similar methods, homologous human (H19230 and R19524) and mouse (AA286626) EST clones were also identified and partially sequenced.

A second T-type calcium channel, termed α1H, was isolated by screening a human heart cDNA library with a fragment of the α1G sequence. The cDNA sequence of α1H is set forth at SEQ ID NO:4, and the deduced amino acid sequence is set forth at SEQ ID NO:3. Also, figure 2 sets forth these sequences and indicates the subunits and putative transmembrane domains of the protein.

The α1G and α1H clones were compared to each other and a known calcium channel (α1E) to investigate the conservation of protein structure and function. The comparison indicates that the α1G and α1H amino acid sequences within the putative membrane-spanning domains are 91% identical to each other, while the α1G and α1H sequences are only 39% identical to the α1E clone. Within the critical IVS4 region, the α1G and α1H proteins are 100% identical, while each is only 44% identical to the α1E clone

Figure 3 indicates this conservation between the proteins. The conservation of charged residues, particularly in the S4 domains, is consistent with the role of the α1G and α1H proteins as ion channels. However, two of the glutamates associated with ion

specificity in other calcium channels have been replaced with aspartate, suggesting altered ion selectivity. Strikingly, both α1G and α1H display only low homology to sequences linking the membrane-spanning regions within each domain, and even less homology between the intracellular loops linking domains. Notably, neither α1G nor α1H possesses sequences known to bind β subunits or Ca^{2+} ions.

EXAMPLE 2

This example demonstrates that the two cloned putative T-type calcium channels exhibit T-type current-voltage relationships.

The α1G and α1H proteins were produced in *Xenopus laevis* oocytes by linearizing the DNA vectors containing the coding sequences, and translating the coding sequences *in vitro* by standard methods. Oocytes were then injected with the capped RNA as described.

Current traces were elicited by depolarizing voltage clamp pulses of the membranes of cells. Figures 4A-5E depict data obtained from these experiments using cells injected with α1G and α1H (Figure 4A and 4B, respectively) and α1E (Figure 4C), as well as undifferentiated NIE-115 cells (Figure 4D), which exhibit classic T-type calcium current (Shuba et al., *J. Physiol. (London), 443*, 25-44 (1991)). These data indicate that cells expressing α1G and α1H (Figure 4A) exhibit T-type calcium current, while oocytes expressing α1E (Figure 4C) as well as uninjected oocytes (Figure 6A) do not.

Current voltage curves were developed using cells injected with α1G, α1H, and α1E, as well as undifferentiated NIE-115 cells. Figure 5A depicts such data generated in a 10 mM Ba^{2+} test solution. These data were transformed into conductance (Figure 5B) and fit with a Boltzman equation to determine the midpoint of activation ($V_{0.5}$). Both NIE-115 cells and α1G currents exhibited low gating potentials (-41 mV ± 1 mV, n=10 and -38 ± 1 mV n=8, respectively), while α1E required significantly more positive potentials to open (-2.6 mV ± .4 mV, n=3).

To compare the characteristics with published values (Huguenard, *Ann. Rev. Physiol., 58*, 329-48 (1996)), the α1G current was recorded at varying concentrations of Ba^{2+}. As indicated in Figure 5C, in solutions containing 2 mM Ba^{2+}, $V_{0.5}$ was -46.5 mV, and the slope factor (*k*) was 6.6 (n=7). However, when the Ba^{2+} concentration was 40 mM, $V_{0.5}$ was recorded at -21 mV, presumably due to the results of barium on surface charge screening (see, e.g., Wilson et al., *J. Membrane Biol., 72*, 117-30 (1983)). Similar values were recorded for α1H.

These results indicate that α1G and α1H are low-voltage activated calcium channels.

EXAMPLE 3

This example demonstrates that the cloned putative T-type calcium channels exhibit T-type kinetics.

To measure activation and inactivation kinetics, oocytes injected with α1E or α1G were pulsed with -20 mV current in 40 mM Ba^{2+}. Data representing the average of five sweeps recorded at 2 kHz and filtered at 1 kHz are presented in Figure 6A. The time constants for α1G inactivation and activation were determined by fitting the data with exponentials. These data are depicted in Figure 6B. These values correspond with the kinetics of the T-type calcium current.

EXAMPLE 4

This example demonstrates that the cloned putative T-type calcium channels exhibit T-type tail deactivation current.

Tail current was measured by prepulsing the cells expressing α1G (oocytes) and α1H (HEK 293 cells) at -90 mV followed by periodic pulses at -10 mV or a pulse at -50 mV. The recordings in Figures 7A and 7B indicate that the current elicited at -50 mV follows the current measured at -10 mV. These data confirm that the decline in current is due to inactivation, rather than activation of a contaminating outward current.

The voltage-dependence of tail current was measured at varying test potentials. Data representing such studies are presented in Figures 7C and 7D, respectively. The data were fit with a single exponential and plotted as a function of depolarization potential (Figures 7E and 7F, respectively). These results demonstrate that the tail currents for the two cloned calcium channels, α1G and α1H, are voltage-dependent, consistent with known T-type calcium tail currents.

EXAMPLE 5

This example demonstrates that the cloned putative T-type calcium channels exhibit T-type voltage dependent inactivation.

To measure inactivation, oocytes expressing α1G or α1E were subjected to 200 ms pre-pulses at various potentials followed by a test pulse to -20 mV. The results of these assays are depicted in Figure 8A.

The data for the 200 ms prepulse experiments were averaged and plotted as a function of prepulse potential (Figure 8B, n= 2 or 4), with a control defined as the current measured after a prepulse of -100 mV.

To approximate steady state conditions, similar experiments were conducted using 10 s prepulses. Inactivation of α1G occurred as sub-threshold potentials and displayed a

steep voltage dependence ($V_{0.5}$ = -50.0 ± 0.2 mV, k = -3.2 ± 0.2, n=5), while inactivation of cloned α1E exhibited more positive potential and weaker voltage dependence ($V_{0.5}$ = -30.0 ± 0.4 mV, k = -9.4 ± 0.3, n=6). These data are depicted in Figure 8C.

EXAMPLE 6

This example demonstrates that the cloned putative T-type calcium channels exhibit T-type single channel conductance.

Measurement of single channel conductance is a function of the low probability of channel opening at negative potentials when the driving force is large. Thus, single channel conductance was measured similarly for measurements of tail currents to enhance channel opening at negative potentials. Single channels were measured with standard depolarizing bath and pipette (115 mM BaCl2, 1 mM EGTA, and 10 mM HEPES, pH 7.4) solutions (Lacerda et al., *Biophys. J.*, 66, 1833-43 (1994)). Data were analyzed with TRANSIT (VanDongan, *Biophys J.*, 70, 1303-15 (1996)). Single channel amplitudes were measured be averaging the values obtained from Gaussian fits to all-points histograms of traces with openings, selected openings, or amplitude histograms of idealized openings. It has been reported that some oocytes contain a native 9 pS channel. These endogenous channels can be distinguished by their 2-fold larger current amplitudes at the potentials tested (e.g., -20 mV, i = 0.8 for endogenous channels as opposed to 0.4 pA for α1G). However, such endogenous channels were not detected either at the whole cell or single channel level in the oocytes tested.

Data were recorded from a patch in oocytes expressing large (>500 nA) α1G currents using a 5 ms step to -20 mV followed by repolarization at potentials indicated in Figure 9A. Data were acquired at 10 kHz and filtered at 2 kHz online and again at 1 kHz off-line. The numbers on the right in Figure 9A indicate the numbers of channels open at any given time.

Current through the main open state of each open channel was measured at each potential and plotted against each test potential. These data are depicted in Figure 9C. Single channel conductance for seven patches were averaged. The average slope conductance of the α1G channel was measured at 7.5 ± 1.5 pS, which corresponds with the reported values for T-type calcium channels (Hugenard, *Ann. Rev. Phsysiol., 58*, 329-48 (1996)).

An ensemble current from 100 sweeps at a -40 mV test current was prepared from the idealized data and fit with a single exponential (τ = 8 ms). This ensemble current is depicted in Figure 9B. This ensemble current exhibits decay kinetics similar to that observed in the macroscopic current measured above (see Figure 7A).

These results indicate that the cloned α1G protein exhibits T-type single-channel conductance.

EXAMPLE 7

This example demonstrates that a cloned T-type calcium channel can be used for identifying a drug which affects T-type calcium channels.

Cells were subjected to treatment as indicated above in Example 2, except that an experimental group of cells were exposed to a solution containing 100 μM mibefradel, a known inhibitor of T-type calcium current. As depicted in Figure 10A, the presence of mibefradel almost completely abolished T-type current in cells expressing α1G. Cells were similarly treated using various concentrations of mibefradel to determine a dose-response relationship. These results, depicted in Figure 10B, demonstrate that 50% inhibition was achieved at a mibefradel concentration of 23 μM.

All of the references cited herein, including patents, patent applications, and publications, are hereby incorporated in their entireties by reference.

While this invention has been described with an emphasis upon preferred embodiments, it will be obvious to those of ordinary skill in the art that variations of the preferred embodiments may be used and that it is intended that the invention may be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications encompassed within the spirit and scope of the invention as defined by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Edward Perez-Reyes

(B) STREET: 2160 S. First Avenue, Building 102, Room 4669

(C) CITY: Maywood

(D) STATE: IL

(E) COUNTRY: US

(F) POSTAL CODE (ZIP): 60153

(G) TELEPHONE: (708) 216-6305

(H) TELEFAX:

(I) TELEX:

(ii) TITLE OF INVENTION: P-Type Voltage-Gated Calcium Channels and Method of Using Same

(iii) NUMBER OF SEQUENCES: 5

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6096 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ATG ACC GAG GGC GCA CGG GCC GCC GAC GAG GTC CGG GTG CCC CTG GGG      48
Met Thr Glu Gly Ala Arg Ala Ala Asp Glu Val Arg Val Pro Leu Gly
 1               5                  10                  15

CGC CGC CCC TGG CCC TGC GGC GTT GGT GGG GGC GTC CCC GGA GAG CCC      96
Arg Arg Pro Trp Pro Cys Gly Val Gly Gly Gly Val Pro Gly Glu Pro
            20                  25                  30

CGG GGC GCC GGG ACG CGA GGC GGA GGG GGG TTC GAG CTC GGC GTG TCA     144
Arg Gly Ala Gly Thr Arg Gly Gly Gly Gly Phe Glu Leu Gly Val Ser
        35                  40                  45

CCC TCC GAG AGC CCG GCG GCC GAG CGC TGC GCG GAG CTG GGT GCC GAC     192
Pro Ser Glu Ser Pro Ala Ala Glu Arg Cys Ala Glu Leu Gly Ala Asp
    50                  55                  60

GAG GAG CAG CGC GTC CCG TAC CCG GCC TTG GCG GCC ACG GTC TTC TTC     240
Glu Glu Gln Arg Val Pro Tyr Pro Ala Leu Ala Ala Thr Val Phe Phe
65                  70                  75                  80

TGC CTC GGT CAG ACC ACG CGG CCG CGC AGC TGG TCC GTC CGG CTG GTC     288
Cys Leu Gly Gln Thr Thr Arg Pro Arg Ser Trp Ser Val Arg Leu Val
                85                  90                  95

TGC AAC CCA TGG TTC GAG CAC GTG AGC ATG CTG GTA ATC ATG CTC AAC     336
Cys Asn Pro Trp Phe Glu His Val Ser Met Leu Val Ile Met Leu Asn
            100                 105                 110

TGC GTG ACC CTG GGC ATG TTC CGG CCC TGT GAG GAC GTT GAG TGC GGC     384
Cys Val Thr Leu Gly Met Phe Arg Pro Cys Glu Asp Val Glu Cys Gly
        115                 120                 125
```

```
TCC GAG CGC TGC AAC ATC CTG GAG GCC TTT GAC GCC TTC ATT TTC GCC        432
Ser Glu Arg Cys Asn Ile Leu Glu Ala Phe Asp Ala Phe Ile Phe Ala
    130                 135                 140

TTT TTT GCG GTG GAG ATG GTC ATC AAG ATG GTG GCC TTG GGG CTG TTC        480
Phe Phe Ala Val Glu Met Val Ile Lys Met Val Ala Leu Gly Leu Phe
145                 150                 155                 160

GGG CAG AAG TGT TAC CTG GGT GAC ACG TGG AAC AGG CTG GAT TTC TTC        528
Gly Gln Lys Cys Tyr Leu Gly Asp Thr Trp Asn Arg Leu Asp Phe Phe
                165                 170                 175

ATC GTC GTG GCG GGC ATG ATG GAG TAC TCG TTG GAC GGA CAC AAC GTG        576
Ile Val Val Ala Gly Met Met Glu Tyr Ser Leu Asp Gly His Asn Val
            180                 185                 190

AGC CTC TCG GCT ATC AGG ACC GTG CGG GTG CTG CGG CCC CTC CGC GCC        624
Ser Leu Ser Ala Ile Arg Thr Val Arg Val Leu Arg Pro Leu Arg Ala
        195                 200                 205

ATC AAC CGC GTG CCT AGC ATG CGG ATC CTG GTC ACT CTG CTG CTG GAT        672
Ile Asn Arg Val Pro Ser Met Arg Ile Leu Val Thr Leu Leu Leu Asp
    210                 215                 220

ACG CTG CCC ATG CTC GGG AAC GTC CTT CTG CTG TGC TTC TTC GTC TTC        720
Thr Leu Pro Met Leu Gly Asn Val Leu Leu Leu Cys Phe Phe Val Phe
225                 230                 235                 240

TTC ATT TTC GGC ATC GTT GGC GTC CAG CTC TGG GCT GGC CTC CTG CGG        768
Phe Ile Phe Gly Ile Val Gly Val Gln Leu Trp Ala Gly Leu Leu Arg
                245                 250                 255

AAC CGC TGC TTC CTG GAC AGT GCC TTT GTC AGG AAC AAC AAC CTG ACC        816
Asn Arg Cys Phe Leu Asp Ser Ala Phe Val Arg Asn Asn Asn Leu Thr
            260                 265                 270

TTC CTG CGG CCG TAC TAC CAG ACG GAG GAG GGC GAG GAG AAC CCG TTC        864
```

```
Phe Leu Arg Pro Tyr Tyr Gln Thr Glu Glu Gly Glu Glu Asn Pro Phe
        275                 280                 285

ATC TGC TCC TCA CGC CGA GAC AAC GGC ATG CAG AAG TGC TCG CAC ATC      912
Ile Cys Ser Ser Arg Arg Asp Asn Gly Met Gln Lys Cys Ser His Ile
    290                 295                 300

CCC GGC CGC CGC GAC GTG CGC ATG CCC TGC ACC CTG GGC TGG GAG GCC      960
Pro Gly Arg Arg Asp Val Arg Met Pro Cys Thr Leu Gly Trp Glu Ala
305                 310                 315                 320

TAC ACG CAG CCG CAG GCC GAG GGG GTG GGC GCT GCA CGC AAC GCC TGC     1008
Tyr Thr Gln Pro Gln Ala Glu Gly Val Gly Ala Ala Arg Asn Ala Cys
                325                 330                 335

ATC AAC TGG AAC CAG TAC TAC AAC GTG TGC CGC TCG GGT GAC TCC AAC     1056
Ile Asn Trp Asn Gln Tyr Tyr Asn Val Cys Arg Ser Gly Asp Ser Asn
            340                 345                 350

CCC CAC AAC GGT GCC ATC AAC TTC GAC AAC ACC TGC TAC GCC TGG ATT     1104
Pro His Asn Gly Ala Ile Asn Phe Asp Asn Thr Cys Tyr Ala Trp Ile
        355                 360                 365

GCC ATC TTC CAG GTG ATC ACG CTG GAA GGC TGG GTG GAC ATC ATG TAC     1152
Ala Ile Phe Gln Val Ile Thr Leu Glu Gly Trp Val Asp Ile Met Tyr
    370                 375                 380

TAC GTC ATG GAC GCC CAC TCA TTC TAC AAC TTC ATC TAT TTC ATC CTG     1200
Tyr Val Met Asp Ala His Ser Phe Tyr Asn Phe Ile Tyr Phe Ile Leu
385                 390                 395                 400

CTC ATC ATC GTG GGC TCC TTC TTC ATG ATC AAC CTG TGC CTG GTG GTG     1248
Leu Ile Ile Val Gly Ser Phe Phe Met Ile Asn Leu Cys Leu Val Val
                405                 410                 415

ATT GCC ACG CAG TTC TCG GAG ACG AAG CAG CGG GAG AGT CAG CTG ATG     1296
Ile Ala Thr Gln Phe Ser Glu Thr Lys Gln Arg Glu Ser Gln Leu Met
```

```
                420                 425                 430

CGG GAG CAG CGG GCA CGC CAC CTG TCC AAC GAC AGC ACG CTG GCC AGC      1344
Arg Glu Gln Arg Ala Arg His Leu Ser Asn Asp Ser Thr Leu Ala Ser
        435                 440                 445

TTC TCC GAG CCT GGC AGC TGC TAC GAA GAG CTG CTG AAG TAC GTG GGC      1392
Phe Ser Glu Pro Gly Ser Cys Tyr Glu Glu Leu Leu Lys Tyr Val Gly
    450                 455                 460

CAC ATA TTC CGC AAG GTC AAG CGG CAG CTT GCG CCT CTA CGC CCG CTG      1440
His Ile Phe Arg Lys Val Lys Arg Gln Leu Ala Pro Leu Arg Pro Leu
465                 470                 475                 480

GCA GAG CCG TGG CGC AAG AAG GTG GAC CCC AGT GCT GTG CAA GGC CAG      1488
Ala Glu Pro Trp Arg Lys Lys Val Asp Pro Ser Ala Val Gln Gly Gln
                485                 490                 495

GGT CCC GGG CAC CGC CAG CGC CGG GCA GGC AGG CAC ACA GCC TCG GTG      1536
Gly Pro Gly His Arg Gln Arg Arg Ala Gly Arg His Thr Ala Ser Val
            500                 505                 510

CAC CAC CTG GTC TAC CAC CAC CAT CAC CAC CAC CAC CAC CAC TAC CAT      1584
His His Leu Val Tyr His His His His His His His His His Tyr His
        515                 520                 525

TTC AGC CAT GGC AGC CCC CGC AGG CCC GGC CCC GAG CCA GGC GCC TGC      1632
Phe Ser His Gly Ser Pro Arg Arg Pro Gly Pro Glu Pro Gly Ala Cys
    530                 535                 540

GAC ACC AGG CTG GTC CGA GCT GGC GCG CCC CCC TCG CCA CCT TCC CCA      1680
Asp Thr Arg Leu Val Arg Ala Gly Ala Pro Pro Ser Pro Pro Ser Pro
545                 550                 555                 560

GGC CGC GGA CCC CCC GAC GCA GAG TCT GTG CAC AGC ATC TAC CAT GCC      1728
Gly Arg Gly Pro Pro Asp Ala Glu Ser Val His Ser Ile Tyr His Ala
                565                 570                 575
```

```
GAC TGC CAC ATA GAG GGG CCG CAG GAG AGG GCC CGG GTG GGC ACA TGC      1776
Asp Cys His Ile Glu Gly Pro Gln Glu Arg Ala Arg Val Gly Thr Cys
            580                 585                 590

CGC AGC CAC TGC CGC TGC CAG CCT CAG GCT GGC CAC AGG GCT GGG CAC      1824
Arg Ser His Cys Arg Cys Gln Pro Gln Ala Gly His Arg Ala Gly His
        595                 600                 605

CAT GAA CTA CCC CAC GAT CCT GCC CTC AGG GGT GGG CAG CGG CAA AGG      1872
His Glu Leu Pro His Asp Pro Ala Leu Arg Gly Gly Gln Arg Gln Arg
    610                 615                 620

CAG CAC CAG CCC CGG ACC CAA GGG GAA GTG GGC CGG TGG ACC GCC AGG      1920
Gln His Gln Pro Arg Thr Gln Gly Glu Val Gly Arg Trp Thr Ala Arg
625                 630                 635                 640

CAC CGG GGG CAC GGC CCG TTG AGC TTG AAC AGC CCT GAT CCC TAC GAG      1968
His Arg Gly His Gly Pro Leu Ser Leu Asn Ser Pro Asp Pro Tyr Glu
                645                 650                 655

AAG ATC CCG CAT GTG GCC GGG GAG CAT GGA CTG GCC AGC CCT GGC CAT      2016
Lys Ile Pro His Val Ala Gly Glu His Gly Leu Ala Ser Pro Gly His
            660                 665                 670

CTG TCG GGC CTC AGT GTG CCC TGC CCC CTG CCC AGC CCC CCA GCG GGC      2064
Leu Ser Gly Leu Ser Val Pro Cys Pro Leu Pro Ser Pro Pro Ala Gly
        675                 680                 685

ACA CTG ACC TGT GAG CTG AAG AGC TGC CCG TAC TGC ACC CGT GCC CTG      2112
Thr Leu Thr Cys Glu Leu Lys Ser Cys Pro Tyr Cys Thr Arg Ala Leu
    690                 695                 700

GAG GAC CCG GAG GGT GAG CTC AGC GGC TCG GAA AGT GGA GAC TCA GAT      2160
Glu Asp Pro Glu Gly Glu Leu Ser Gly Ser Glu Ser Gly Asp Ser Asp
705                 710                 715                 720
```

```
GGC CGT GGC GTC TAT GAA TTC ACG CAG GAC GTC CGG CAC GGT GAC CGC      2208
Gly Arg Gly Val Tyr Glu Phe Thr Gln Asp Val Arg His Gly Asp Arg
                725                 730                 735

TGG GAC CCC ACG CGA CCA CCC CGT GCG ACG GAC ACA CCA GGC CCA GGC      2256
Trp Asp Pro Thr Arg Pro Pro Arg Ala Thr Asp Thr Pro Gly Pro Gly
            740                 745                 750

CCA GGC AGC CCC CAG CGG CGG GCA CAG CAG AGG GCA GCC CCG GGC GAG      2304
Pro Gly Ser Pro Gln Arg Arg Ala Gln Gln Arg Ala Ala Pro Gly Glu
        755                 760                 765

CCA GGC TGG ATG GGC CGC CTC TGG GTT ACC TTC AGC GGC AAG CTG CGC      2352
Pro Gly Trp Met Gly Arg Leu Trp Val Thr Phe Ser Gly Lys Leu Arg
    770                 775                 780

CGC ATC GTG GAC AGC AAG TAC TTC AGC CGT GGC ATC ATG ATG GCC ATC      2400
Arg Ile Val Asp Ser Lys Tyr Phe Ser Arg Gly Ile Met Met Ala Ile
785                 790                 795                 800

CTT GTC AAC ACG CTG AGC ATG GGC GTG GAG TAC CAT GAG CAG CCC GAG      2448
Leu Val Asn Thr Leu Ser Met Gly Val Glu Tyr His Glu Gln Pro Glu
                805                 810                 815

GAG CTG ACT AAT GCT CTG GAG ATC AGC AAC ATC GTG TTC ACC AGC ATG      2496
Glu Leu Thr Asn Ala Leu Glu Ile Ser Asn Ile Val Phe Thr Ser Met
            820                 825                 830

TTT GCC CTG GAG ATG CTG CTG AAG CTG CTG CGC GCT GTC CCT CTG GGC      2544
Phe Ala Leu Glu Met Leu Leu Lys Leu Leu Arg Ala Val Pro Leu Gly
        835                 840                 845

TAC ATC CGG AAC CCG TAC AAC ATC TTC GAC GGC ATC ATC GTG GTC ATC      2592
Tyr Ile Arg Asn Pro Tyr Asn Ile Phe Asp Gly Ile Ile Val Val Ile
    850                 855                 860

AGC GTC TGG GAG ATC GTG GGG CAG GCG GAC GGT GGC TTG TCT GTG CTG      2640
```

```
Ser Val Trp Glu Ile Val Gly Gln Ala Asp Gly Gly Leu Ser Val Leu
865                 870                 875                 880

CGC ACC TTC CGG CTG CTG CGT GTG CTG AAG CTG GTG CGC TTT CTG CCA      2688
Arg Thr Phe Arg Leu Leu Arg Val Leu Lys Leu Val Arg Phe Leu Pro
                885                 890                 895

GCC CTG CGG CGC CAG CTC GTG GTG CTG GTG AAG ACC ATG GAC AAC GTG      2736
Ala Leu Arg Arg Gln Leu Val Val Leu Val Lys Thr Met Asp Asn Val
        900                 905                 910

GCT ACC TTC TGC ACG CTG CTC ATG CTC TTC ATT TTC ATC TTC AGC ATC      2784
Ala Thr Phe Cys Thr Leu Leu Met Leu Phe Ile Phe Ile Phe Ser Ile
    915                 920                 925

CTG GGC ATG CAC CTT TTC GGC TGC AAG TTC AGC CTG AAG ACA GAC ACC      2832
Leu Gly Met His Leu Phe Gly Cys Lys Phe Ser Leu Lys Thr Asp Thr
930                 935                 940

GGA GAC ACC GTG CCT GAC AGG AAG AAC TTC GAC TCC CTG CTG TGG GCC      2880
Gly Asp Thr Val Pro Asp Arg Lys Asn Phe Asp Ser Leu Leu Trp Ala
945                 950                 955                 960

ATC GTC ACC GTG TTC CAG ATC CTG ACC CAG GAG GAC TGG AAC GTG GTC      2928
Ile Val Thr Val Phe Gln Ile Leu Thr Gln Glu Asp Trp Asn Val Val
                965                 970                 975

CTG TAC AAC GGC ATG GCC TCC ACC TCC TCC TGG GCC GCC CTC TAC TTC      2976
Leu Tyr Asn Gly Met Ala Ser Thr Ser Ser Trp Ala Ala Leu Tyr Phe
            980                 985                 990

GTG GCC CTC ATG ACC TTC GGC AAC TAT GTG CTC TTC AAC CTG CTG GTG      3024
Val Ala Leu Met Thr Phe Gly Asn Tyr Val Leu Phe Asn Leu Leu Val
        995                 1000                1005

GCC ATC CTC GTG GAG GGC TTC CAG GCG GAG GGC GAT GCC AAC AGA TCC      3072
Ala Ile Leu Val Glu Gly Phe Gln Ala Glu Gly Asp Ala Asn Arg Ser
```

1010 1015 1020

GAC ACG GAC GAG GAC AAG ACG TCG GTC CAC TTC GAG GAG GAC TTC CAC 3120
Asp Thr Asp Glu Asp Lys Thr Ser Val His Phe Glu Glu Asp Phe His
1025 1030 1035 1040

AAG CTC AGA GAA CTC CAG ACC ACA GAG CTG AAG ATG TGT TCC CTG GCC 3168
Lys Leu Arg Glu Leu Gln Thr Thr Glu Leu Lys Met Cys Ser Leu Ala
1045 1050 1055

GTG ACC CCC AAC GGC ACC TGG AGG GAC GAG GCA GCC TGT CCC CTC CCC 3216
Val Thr Pro Asn Gly Thr Trp Arg Asp Glu Ala Ala Cys Pro Leu Pro
1060 1065 1070

TCA TCA TGT GCA CAG CTG CCA CGC CCA TGC CTA CCC CCA AGA GCT CAC 3264
Ser Ser Cys Ala Gln Leu Pro Arg Pro Cys Leu Pro Pro Arg Ala His
1075 1080 1085

CAT TCC TGG ATG CAG CCC CCA GCC TCC CAG ACT CTC GGC GTG GCA GCA 3312
His Ser Trp Met Gln Pro Pro Ala Ser Gln Thr Leu Gly Val Ala Ala
1090 1095 1100

GCA GCT CCG GGG ACC CGC CAC TGG GAG ACC AGA AGC CTC CGG CAG CCT 3360
Ala Ala Pro Gly Thr Arg His Trp Glu Thr Arg Ser Leu Arg Gln Pro
1105 1110 1115 1120

CCG AAG TTC TCC CTG TGC CCC CTG GGG CCC AGT GGC GCC TGG AGC AGC 3408
Pro Lys Phe Ser Leu Cys Pro Leu Gly Pro Ser Gly Ala Trp Ser Ser
1125 1130 1135

CGG CGC TCC AGC TGG AGC AGC CTG GGC CGT GCC CAG CCT CAA GCG CCG 3456
Arg Arg Ser Ser Trp Ser Ser Leu Gly Arg Ala Gln Pro Gln Ala Pro
1140 1145 1150

GCG TGC CAG TGT GGG GAA CGT GAG TCC CTG CTG TCT GGC GAG GGC AAG 3504
Ala Cys Gln Cys Gly Glu Arg Glu Ser Leu Leu Ser Gly Glu Gly Lys
1155 1160 1165

```
GGC AGC ACC GAC GAC GAA GCT GAG GAC GGC AGG GCG CGC TCC GGG CCC      3552
Gly Ser Thr Asp Asp Glu Ala Glu Asp Gly Arg Ala Arg Ser Gly Pro
    1170                1175                1180

CGT GCC ACC CCA CTG CGG CGG GCC GAG TCC CTG GAC CCA CGG CCC CTG      3600
Arg Ala Thr Pro Leu Arg Arg Ala Glu Ser Leu Asp Pro Arg Pro Leu
1185                1190                1195                1200

CGG CGG CCG CCT CCC GCC TAC CAA GTG CGC GAT CGC GAC GGG CAG GTG      3648
Arg Arg Pro Pro Pro Ala Tyr Gln Val Arg Asp Arg Asp Gly Gln Val
                1205                1210                1215

GTG GCC CTG CCC AGC GAC TTC TTC CTG CGC ATC GAC AGC CAC CGT GAG      3696
Val Ala Leu Pro Ser Asp Phe Phe Leu Arg Ile Asp Ser His Arg Glu
            1220                1225                1230

GAT GCA GCC GAG CTT GAC GAC GAC TCG GAG GAC AGC TGC TGC CTC CGC      3744
Asp Ala Ala Glu Leu Asp Asp Asp Ser Glu Asp Ser Cys Cys Leu Arg
        1235                1240                1245

CTG CAT AAA GTG CTG GTG CCC TAC AAG CCC CAG CGG TGC CGG AGC AGG      3792
Leu His Lys Val Leu Val Pro Tyr Lys Pro Gln Arg Cys Arg Ser Arg
    1250                1255                1260

AGG CCT GGG CCC TCT ACC CTC TAC CTC TTC TCC CCA CAG AAC CGG TTC      3840
Arg Pro Gly Pro Ser Thr Leu Tyr Leu Phe Ser Pro Gln Asn Arg Phe
1265                1270                1275                1280

CGC GTC TCC TGC CAG AAG GTC ATC ACA CAC AAG ATG TTT GAT CAC GTG      3888
Arg Val Ser Cys Gln Lys Val Ile Thr His Lys Met Phe Asp His Val
                1285                1290                1295

GTC CTC GTC TTC ATC TTC CTC AAC TGC GTC ACC ATC GCC CTG GAG AGG      3936
Val Leu Val Phe Ile Phe Leu Asn Cys Val Thr Ile Ala Leu Glu Arg
            1300                1305                1310
```

```
CCT GAC ATC GAT CCC GGC AGC ACC GAG CGG GTC TTC CTC AGC GTC TCC      3984
Pro Asp Ile Asp Pro Gly Ser Thr Glu Arg Val Phe Leu Ser Val Ser
        1315                1320                1325

AAT TAC ATC TTC ACG GCC ATC TTC GTG GCG GAG ATG ATG GTG AAG GTG      4032
Asn Tyr Ile Phe Thr Ala Ile Phe Val Ala Glu Met Met Val Lys Val
    1330                1335                1340

GTG GCC CTG GGG CTG CTG TCC GGC GAG CAC GCC TAC CTG CAG AGC AGC      4080
Val Ala Leu Gly Leu Leu Ser Gly Glu His Ala Tyr Leu Gln Ser Ser
1345                1350                1355                1360

TGG AAC CTG CTG GAT GGG CTG CTG GTG CTG GTG TCC CTG GTG GAC ATT      4128
Trp Asn Leu Leu Asp Gly Leu Leu Val Leu Val Ser Leu Val Asp Ile
                1365                1370                1375

GTC GTG GCC ATG GCC TCG GCT GGT GGC GCC AAG ATC CTG GGT GTT CTG      4176
Val Val Ala Met Ala Ser Ala Gly Gly Ala Lys Ile Leu Gly Val Leu
            1380                1385                1390

CGC GTG CTG CGT CTG CTG CGG ACC CTG CGG CCT CTG AGG GTC ATC AGC      4224
Arg Val Leu Arg Leu Leu Arg Thr Leu Arg Pro Leu Arg Val Ile Ser
        1395                1400                1405

CGG CCC CGG CTC AAG CTG GTG GTG GAG ACG CTG ATA TCA TCA CTC AGG      4272
Arg Pro Arg Leu Lys Leu Val Val Glu Thr Leu Ile Ser Ser Leu Arg
    1410                1415                1420

CCC ATT GGG AAC ATC GTC CTC ATC TGC TGC GCC TTC TTC ATC ATT TTT      4320
Pro Ile Gly Asn Ile Val Leu Ile Cys Cys Ala Phe Phe Ile Ile Phe
1425                1430                1435                1440

GGC ATT TTG GGT GTG CAG CTC TTC AAA GGG AAG TTC TAC TAC TGC GAG      4368
Gly Ile Leu Gly Val Gln Leu Phe Lys Gly Lys Phe Tyr Tyr Cys Glu
                1445                1450                1455

GGC CCC GAC ACC AGG AAC ATC TCC ACC AAG GCA CAG TGC CGG GCC GCC      4416
```

```
Gly Pro Asp Thr Arg Asn Ile Ser Thr Lys Ala Gln Cys Arg Ala Ala
            1460                1465                1470

CAC TAC CGC TGG GTG CGA CGC AAG TAC AAC TTC GAC AAC CTG GGC CAG      4464
His Tyr Arg Trp Val Arg Arg Lys Tyr Asn Phe Asp Asn Leu Gly Gln
        1475                1480                1485

GCC CTG ATG TCG CTG TTC GTG CTG TCA TCC AAG GAT GGA TGG GTG AAC      4512
Ala Leu Met Ser Leu Phe Val Leu Ser Ser Lys Asp Gly Trp Val Asn
    1490                1495                1500

ATC ATG TAC GAC GGG CTG GAT GCC GTG GGT GTC GAC CAG CAG CCT GTG      4560
Ile Met Tyr Asp Gly Leu Asp Ala Val Gly Val Asp Gln Gln Pro Val
1505                1510                1515                1520

CAG AAC CAC AAC CCC TGG ATG CTG CTG TAC TTC ATC TCC TTC CTC TGC      4608
Gln Asn His Asn Pro Trp Met Leu Leu Tyr Phe Ile Ser Phe Leu Cys
                1525                1530                1535

TAC ATC GTC AGC TTC TTC GTG CTC AAC ATG TTC GTG GGC GTC GTG GTC      4656
Tyr Ile Val Ser Phe Phe Val Leu Asn Met Phe Val Gly Val Val Val
            1540                1545                1550

GAG AAC TTC CAC AAG TGC CGG CCG CAC CAG GAG GCG GAG GAG GCG CGG      4704
Glu Asn Phe His Lys Cys Arg Pro His Gln Glu Ala Glu Glu Ala Arg
        1555                1560                1565

CGG CGA GAG GAG AAG CGG CTG CGG CGC CTA GAG AGG AGG CGC AGG AGC      4752
Arg Arg Glu Glu Lys Arg Leu Arg Arg Leu Glu Arg Arg Arg Arg Ser
    1570                1575                1580

ACT TTC CCC AGC CCA GAG GCC CAG CGC CGG CCC TAC TAT GCC GAC TAC      4800
Thr Phe Pro Ser Pro Glu Ala Gln Arg Arg Pro Tyr Tyr Ala Asp Tyr
1585                1590                1595                1600

TCG CCC ACG CGC CGC CGC TCC ATT CAC TCG CTG TGC ACC AGC CAC TAT      4848
Ser Pro Thr Arg Arg Arg Ser Ile His Ser Leu Cys Thr Ser His Tyr
```

```
                1605                1610                1615

CTC GAC CTC TTC ATC ACC TTC ATC ATC TGT GTC AAC GTC ATC ACC ATG      4896
Leu Asp Leu Phe Ile Thr Phe Ile Ile Cys Val Asn Val Ile Thr Met
            1620                1625                1630

TCC ATG GAG CAC TAT AAC CAA CCC AAG TCG CTG GAC GAG GCC CTC AAG      4944
Ser Met Glu His Tyr Asn Gln Pro Lys Ser Leu Asp Glu Ala Leu Lys
        1635                1640                1645

TAC TGC AAC TAC GTC TTC ACC ATC GTG TTT GTC TTC GAG GCT GCA CTG      4992
Tyr Cys Asn Tyr Val Phe Thr Ile Val Phe Val Phe Glu Ala Ala Leu
    1650                1655                1660

AAG CTG GTA GCA TTT GGG TTC CGT CGG TTC TTC AAG GAC AGG TGG AAC      5040
Lys Leu Val Ala Phe Gly Phe Arg Arg Phe Phe Lys Asp Arg Trp Asn
1665                1670                1675                1680

CAG CTG GAC CTG GCC ATC GTG CTG CTG TCA CTC ATG GGC ATC ACG CTG      5088
Gln Leu Asp Leu Ala Ile Val Leu Leu Ser Leu Met Gly Ile Thr Leu
                1685                1690                1695

GAG GAG ATA GAG ATG AGC GCC GCG CTG CCC ATC AAC CCC ACC ATC ATC      5136
Glu Glu Ile Glu Met Ser Ala Ala Leu Pro Ile Asn Pro Thr Ile Ile
            1700                1705                1710

CGC ATC ATG CGC GTG CTT CGC ATT GCC CGT GTG CTG AAG CTG CTG AAG      5184
Arg Ile Met Arg Val Leu Arg Ile Ala Arg Val Leu Lys Leu Leu Lys
        1715                1720                1725

ATG GCT ACG GGC ATG CGC GCC CTG CTG GAC ACT GTG GTG CAA GCT CTC      5232
Met Ala Thr Gly Met Arg Ala Leu Leu Asp Thr Val Val Gln Ala Leu
    1730                1735                1740

CCC CAG GTG GGG AAC CTG GGC CTT CTT TTC ATG CTC CTG TTT TTT ATC      5280
Pro Gln Val Gly Asn Leu Gly Leu Leu Phe Met Leu Leu Phe Phe Ile
1745                1750                1755                1760
```

```
TAT CTG AGA TTG GGA GTG GAG CTG TTC GGG AGG CTG GAG TGC AGT GAA      5328
Tyr Leu Arg Leu Gly Val Glu Leu Phe Gly Arg Leu Glu Cys Ser Glu
                1765                1770                1775

GAC AAC CCC TGC GAG GGC CTG AGC AGG CAC GCC ACC TTC AGC AAC TTC      5376
Asp Asn Pro Cys Glu Gly Leu Ser Arg His Ala Thr Phe Ser Asn Phe
            1780                1785                1790

GGC ATG GCC TTC CTC ACG CTG TTC CGC GTG TCC ACG GGG GAC AAC TGG      5424
Gly Met Ala Phe Leu Thr Leu Phe Arg Val Ser Thr Gly Asp Asn Trp
        1795                1800                1805

AAC GGG ATC ATG AAG GAC ACG CTG CGC GAG TGC TCC CGT GAG GAC AAG      5472
Asn Gly Ile Met Lys Asp Thr Leu Arg Glu Cys Ser Arg Glu Asp Lys
    1810                1815                1820

CAC TGC CTG AGC TAC CTG CCG GCC CCG TCG CCC GTC TAC TTC GTG ACC      5520
His Cys Leu Ser Tyr Leu Pro Ala Pro Ser Pro Val Tyr Phe Val Thr
1825                1830                1835                1840

TTC GTG CTG GTG CCC CAG TTC GTG CTG GTG AAC GTG GTG GTG GCC GTG      5568
Phe Val Leu Val Pro Gln Phe Val Leu Val Asn Val Val Val Ala Val
                1845                1850                1855

CTC ATG AAG CAC CTG GAG GAG AGC AAC AAG GAG GCT CGG GAG GAT GCG      5616
Leu Met Lys His Leu Glu Glu Ser Asn Lys Glu Ala Arg Glu Asp Ala
            1860                1865                1870

GAG CTG GAC GCC GAG ATC GAG CTG GAG ATG GCG CAG GGC CCC GGG AGT      5664
Glu Leu Asp Ala Glu Ile Glu Leu Glu Met Ala Gln Gly Pro Gly Ser
        1875                1880                1885

GCA CGC CGG GTG GAC GCG GAC AGG CCT CCC TTG CCC CAG GAG AGT CCG      5712
Ala Arg Arg Val Asp Ala Asp Arg Pro Pro Leu Pro Gln Glu Ser Pro
    1890                1895                1900
```

```
GCG CCA GGG ACG CCC CAA ACC TGG TTG CAC GCA AGG TGT CCG TGT CCA      5760
Ala Pro Gly Thr Pro Gln Thr Trp Leu His Ala Arg Cys Pro Cys Pro
1905                1910                1915                1920

GGA TCT CTC GCT GCC CAA CGA CAG CTA CAT GTT CAG GCC CGT GGT GCC      5808
Gly Ser Leu Ala Ala Gln Arg Gln Leu His Val Gln Ala Arg Gly Ala
                1925                1930                1935

TGC CTC GGC GCC CCG GGC CCG CCC GCT GCA GGA GGT GGA GAT GGA GAC      5856
Cys Leu Gly Ala Pro Gly Pro Pro Ala Ala Gly Gly Gly Asp Gly Asp
            1940                1945                1950

CTA TGG GGC CGG CAC CCC CTT GGA GTC CTG TGC CAT CCC ATC CAG ATC      5904
Leu Trp Gly Arg His Pro Leu Gly Val Leu Cys His Pro Ile Gln Ile
        1955                1960                1965

CCA TTG GCT GTG TCG AAC CCA GCC AGG AGC GGC GAG CCC CTC CAC GCC      5952
Pro Leu Ala Val Ser Asn Pro Ala Arg Ser Gly Glu Pro Leu His Ala
    1970                1975                1980

CTG TCC CCT CGG GGC ACA GCC GCT CCC CCA GTC TCA GCC GGC TGC TCT      6000
Leu Ser Pro Arg Gly Thr Ala Ala Pro Pro Val Ser Ala Gly Cys Ser
1985                1990                1995                2000

GCA GAC AGG AGG CTG TGC ACA CCG ATT CCT TGG AAG GGA AGA TTG ACA      6048
Ala Asp Arg Arg Leu Cys Thr Pro Ile Pro Trp Lys Gly Arg Leu Thr
                2005                2010                2015

GCC CTA GGG ACA CCC TGG ATC CTG CAG AGC CTG GTG AGA AAC CCC CGG      6096
Ala Leu Gly Thr Pro Trp Ile Leu Gln Ser Leu Val Arg Asn Pro Arg
            2020                2025                2030
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7720 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GGCGGTGACC GCGCCGCCCG GCGATGCCCG CGGGGACGCC GCCGGCCAGC AGAGCAGGTG 60

CTGCCGGCCG CCACCATGAC CGAGGGCGCA CGGGCCGCCG ACGAGGTCCG GGTGCCCCTG 120

GGGCGCCGCC CCTGGCCCTG CGGCGTTGGC GGGGGCGTCC CCGGAGAGCC CCGGGGCGCC 180

GGGACGCGAG GCGGAGGGGG GTTCGAGCTC GGCGTGTCAC CCTCCGAGAG CCCGGCGGCC 240

GAGCGCTGCG CGGAGCTGGG TGCCGACGAG GAGCAGCGCG TCCCGTACCC GGCCTTGGCG 300

GCCACGGTCT TCTTCTGCCT CGGTCAGACC ACGCGGCCGC GCAGCTGGTC CGTCCGGCTG 360

GTCTGCAACC CATGGTTCGA GCACGTGAGC ATGCTGGTAA TCATGCTCAA CTGCGTGACC 420

CTGGGCATGT TCCGGCCCTG TGAGGACGTT GAGTGCGGCT CCGAGCGCTG CAACATCCTG 480

GAGGCCTTTG ACGCCTTCAT TTTCGCCTTT TTTGCGGTGG AGATGGTCAT CAAGATGGTG 540

GCCTTGGGGC TGTTCGGGCA GAAGTGTTAC CTGGGTGACA CGTGGAACAG GCTGGATTTC 600

TTCATCGTCG TGGCGGGCAT GATGGAGTAC TCGTTGGACG GACACAACGT GAGCCTCTCG 660

GCTATCAGGA CCGTGCGGGT GCTGCGGCCC CTCCGCGCCA TCAACCGCGT GCCTAGCATG 720

CGGATCCTGG TCACTCTGCT GCTGGATACG CTGCCCATGC TCGGGAACGT CCTTCTGCTG 780

TGCTTCTTCG TCTTCTTCAT TTTCGGCATC GTTGGCGTCC AGCTCTGGGC TGGCCTCCTG 840

```
CGGAACCGCT GCTTCCTGGA CAGTGCCTTT GTCAGGAACA ACAACCTGAC CTTCCTGCGG     900

CCGTACTACC AGACGGAGGA GGGCGAGGAG AACCCGTTCA TCTGCTCCTC ACGCCGAGAC     960

AACGGCATGC AGAAGTGCTC GCACATCCCC GGCCGCCGCG ACGTGCGCAT GCCCTGCACC    1020

CTGGGCTGGG AGGCCTACAC GCAGCCGCAG GCCGAGGGGG TGGGCGCTGC ACGCAACGCC    1080

TGCATCAACT GGAACCAGTA CTACAACGTG TGCCGCTCGG GTGACTCCAA CCCCCACAAC    1140

GGTGCCATCA ACTTCGACAA CACCTGCTAC GCCTGGATTG CCATCTTCCA GGTGATCACG    1200

CTGGAAGGCT GGGTGGACAT CATGTACTAC GTCATGGACG CCCACTCATT CTACAACTTC    1260

ATCTATTTCA TCCTGCTCAT CATCGTGGGC TCCTTCTTCA TGATCAACCT GTGCCTGGTG    1320

GTGATTGCCA CGCAGTTCTC GGAGACGAAG CAGCGGGAGA GTCAGCTGAT GCGGGAGCAG    1380

CGGGCACGCC ACCTGTCCAA CGACAGCACG CTGGCCAGCT TCTCCGAGCC TGGCAGCTGC    1440

TACGAAGAGC TGCTGAAGTA CGTGGGCCAC ATATTCCGCA AGGTCAAGCG GCAGCTTGCG    1500

CCTCTACGCC CGCTGGCAGA GCCGTGGCGC AAGAAGGTGG ACCCCAGTGC TGTGCAAGGC    1560

CAGGGTCCCG GGCACCGCCA GCGCCGGGCA GGCAGGCACA CAGCCTCGGT GCACCACCTG    1620

GTCTACCACC ACCATCACCA CCACCACCAC CACTACCATT TCAGCCATGG CAGCCCCCGC    1680

AGGCCCGGCC CCGAGCCAGG CGCCTGCGAC ACCAGGCTGG TCCGAGCTGG CGCGCCCCCC    1740

TCGCCACCTT CCCCAGGCCG CGGACCCCCC GACGCAGAGT CTGTGCACAG CATCTACCAT    1800

GCCGACTGCC ACATAGAGGG GCCGCAGGAG AGGGCCCGGG TGGGCACATG CCGCAGCCAC    1860

TGCCGCTGCC AGCCTCAGGC TGGCCACAGG GCTGGGCACC ATGAACTACC CCACGATCCT    1920

GCCCTCAGGG GTGGGCAGCG GCAAAGGCAG CACCAGCCCC GGACCCAAGG GGAAGTGGGC    1980
```

CGGTGGACCG CCAGGCACCG GGGGCACGGC CCGTTGAGCT TGAACAGCCC TGATCCCTAC 2040

GAGAAGATCC CGCATGTGGC CGGGGAGCAT GGACTGGCCA GCCCTGGCCA TCTGTCGGGC 2100

CTCAGTGTGC CCTGCCCCCT GCCCAGCCCC CCAGCGGGCA CACTGACCTG TGAGCTGAAG 2160

AGCTGCCCGT ACTGCACCCG TGCCCTGGAG GACCCGGAGG GTGAGCTCAG CGGCTCGGAA 2220

AGTGGAGACT CAGATGGCCG TGGCGTCTAT GAATTCACGC AGGACGTCCG GCACGGTGAC 2280

CGCTGGGACC CCACGCGACC ACCCCGTGCG ACGGACACAC CAGGCCCAGG CCCAGGCAGC 2340

CCCCAGCGGC GGGCACAGCA GAGGGCAGCC CCGGGCGAGC CAGGCTGGAT GGGCCGCCTC 2400

TGGGTTACCT TCAGCGGCAA GCTGCGCCGC ATCGTGGACA GCAAGTACTT CAGCCGTGGC 2460

ATCATGATGG CCATCCTTGT CAACACGCTG AGCATGGGCG TGGAGTACCA TGAGCAGCCC 2520

GAGGAGCTGA CTAATGCTCT GGAGATCAGC AACATCGTGT TCACCAGCAT GTTTGCCCTG 2580

GAGATGCTGC TGAAGCTGCT GCGCGCTGTC CCTCTGGGCT ACATCCGGAA CCCGTACAAC 2640

ATCTTCGACG GCATCATCGT GGTCATCAGC GTCTGGGAGA TCGTGGGGCA GGCGGACGGT 2700

GGCTTGTCTG TGCTGCGCAC CTTCCGGCTG CTGCGTGTGC TGAAGCTGGT GCGCTTTCTG 2760

CCAGCCCTGC GGCGCCAGCT CGTGGTGCTG GTGAAGACCA TGGACAACGT GGCTACCTTC 2820

TGCACGCTGC TCATGCTCTT CATTTTCATC TTCAGCATCC TGGGCATGCA CCTTTTCGGC 2880

TGCAAGTTCA GCCTGAAGAC AGACACCGGA GACACCGTGC CTGACAGGAA GAACTTCGAC 2940

TCCCTGCTGT GGGCCATCGT CACCGTGTTC CAGATCCTGA CCCAGGAGGA CTGGAACGTG 3000

GTCCTGTACA ACGGCATGGC CTCCACCTCC TCCTGGGCCG CCCTCTACTT CGTGGCCCTC 3060

ATGACCTTCG GCAACTATGT GCTCTTCAAC CTGCTGGTGG CCATCCTCGT GGAGGGCTTC 3120

CAGGCGGAGG GCGATGCCAA CAGATCCGAC ACGGACGAGG ACAAGACGTC GGTCCACTTC 3180

GAGGAGGACT TCCACAAGCT CAGAGAACTC CAGACCACAG AGCTGAAGAT GTGTTCCCTG 3240

GCCGTGACCC CCAACGGCAC CTGGAGGGAC GAGGCAGCCT GTCCCCTCCC CTCATCATGT 3300

GCACAGCTGC CACGCCCATG CCTACCCCCA AGAGCTCACC ATTCCTGGAT GCAGCCCCCA 3360

GCCTCCCAGA CTCTCGGCGT GGCAGCAGCA GCTCCGGGGA CCCGCCACTG GGAGACCAGA 3420

AGCCTCCGGC AGCCTCCGAA GTTCTCCCTG TGCCCCCTGG GGCCCAGTGG CGCCTGGAGC 3480

AGCCGGCGCT CCAGCTGGAG CAGCCTGGGC CGTGCCCAGC CTCAAGCGCC GGCGTGCCAG 3540

TGTGGGGAAC GTGAGTCCCT GCTGTCTGGC GAGGGCAAGG GCAGCACCGA CGACGAAGCT 3600

GAGGACGGCA GGGCGCGCTC CGGGCCCCGT GCCACCCCAC TGCGGCGGGC CGAGTCCCTG 3660

GACCCACGGC CCCTGCGGCG GCCGCCTCCC GCCTACCAAG TGCGCGATCG CGACGGGCAG 3720

GTGGTGGCCC TGCCCAGCGA CTTCTTCCTG CGCATCGACA GCCACCGTGA GGATGCAGCC 3780

GAGCTTGACG ACGACTCGGA GGACAGCTGC TGCCTCCGCC TGCATAAAGT GCTGGTGCCC 3840

TACAAGCCCC AGCGGTGCCG GAGCAGGAGG CCTGGGCCCT CTACCCTCTA CCTCTTCTCC 3900

CCACAGAACC GGTTCCGCGT CTCCTGCCAG AAGGTCATCA CACACAAGAT GTTTGATCAC 3960

GTGGTCCTCG TCTTCATCTT CCTCAACTGC GTCACCATCG CCCTGGAGAG GCCTGACATT 4020

GATCCCGGCA GCACCGAGCG GGTCTTCCTC AGCGTCTCCA ACTACATCTT CACGGCCATC 4080

TTCGTGGCGG AGATGATGGT GAAGGTGGTG GCCCTGGGGC TGCTGTCCGG CGAGCACGCC 4140

TACCTGCAGA GCAGCTGGAA CCTGCTGGAT GGGCTGCTGG TGCTGGTGTC CCTGGTGGAC 4200

ATTGTCGTGG CCATGGCCTC GGCTGGTGGC GCCAAGATCC TGGGTGTTCT GCGCGTGCTG 4260

CGTCTGCTGC GGACCCTGCG GCCTCTGAGG GTCATCAGCC GGCCCCGGCT CAAGCTGGTG 4320

GTGGAGACGC TGATATCATC ACTCAGGCCC ATTGGGAACA TCGTCCTCAT CTGCTGCGCC 4380

TTCTTCATCA TTTTTGGCAT TTTGGGTGTG CAGCTCTTCA AAGGGAAGTT CTACTACTGC 4440

GAGGGCCCCG ACACCAGGAA CATCTCCACC AAGGCACAGT GCCGGGCCGC CCACTACCGC 4500

TGGGTGCGAC GCAAGTACAA CTTCGACAAC CTGGGCCAGG CCCTGATGTC GCTGTTCGTG 4560

CTGTCATCCA AGGATGGATG GGTGAACATC ATGTACGACG GGCTGGATGC CGTGGGTGTC 4620

GACCAGCAGC CTGTGCAGAA CCACAACCCC TGGATGCTGC TGTACTTCAT CTCCTTCCTC 4680

TGCTACATCG TCAGCTTCTT CGTGCTCAAC ATGTTCGTGG GCGTCGTGGT CGAGAACTTC 4740

CACAAGTGCC GGCCGCACCA GGAGGCGGAG GAGGCGCGGC GGCGAGAGGA GAAGCGGCTG 4800

CGGCGCCTAG AGAGGAGGCG CAGGAGCACT TCCCCCAGCC CAGAGGCCCA GCGCCGGCCC 4860

TACTATGCCG ACTACTCGCC CACGCGCCGC CGCTCCATTC ACTCGCTGTG CACCAGCCAC 4920

TATCTCGACC TCTTCATCAC CTTCATCATC TGTGTCAACG TCATCACCAT GTCCATGGAG 4980

CACTATAACC AACCCAAGTC GCTGGACGAG GCCCTCAAGT ACTGCAACTA CGTCTTCACC 5040

ATCGTGTTTG TCTTCGAGGC TGCACTGAAG CTGGTAGCAT TTGGGTTCCG TCGGTTCTTC 5100

AAGGACAGGT GGAACCAGCT GGACCTGGCC ATCGTGCTGC TGTCACTCAT GGGCATCACG 5160

CTGGAGGAGA TAGAGATGAG CGCCGCGCTG CCCATCAACC CCACCATCAT CCGCATCATG 5220

CGCGTGCTTC GCATTGCCCG TGTGCTGAAG CTGCTGAAGA TGGCTACGGG CATGCGCGCC 5280

CTGCTGGACA CTGTGGTGCA AGCTCTCCCC CAGGTGGGGA ACCTGGGCCT TCTTTTCATG 5340

CTCCTGTTTT TTATCTATCT GAGATTGGGA GTGGAGCTGT TCGGGAGGCT GGAGTGCAGT 5400

GAAGACAACC CCTGCGAGGG CCTGAGCAGG CACGCCACCT TCAGCAACTT CGGCATGGCC 5460

TTCCTCACGC TGTTCCGCGT GTCCACGGGG GACAACTGGA ACGGGATCAT GAAGGACACG 5520

CTGCGCGAGT GCTCCCGTGA GGACAAGCAC TGCCTGAGCT ACCTGCCGGC CCCGTCGCCC 5580

GTCTACTTCG TGACCTTCGT GCTGGTGCCC CAGTTCGTGC TGGTGAACGT GGTGGTGGCC 5640

GTGCTCATGA AGCACCTGGA GGAGAGCAAC AAGGAGGCTC GGGAGGATGC GGAGCTGGAC 5700

GCCGAGATCG AGCTGGAGAT GGCGCAGGGC CCCGGGAGTG CACGCCGGGT GGACGCGGAC 5760

AGGCCTCCCT TGCCCCAGGA GAGTCCGGCG CCAGGGACGC CCCAAACCTG GTTGCACGCA 5820

AGGTGTCCGT GTCCAGGATC TCTCGCTGCC CAACGACAGC TACATGTTCA GGCCCGTGGT 5880

GCCTGCCTCG GCGCCCCGGG CCCGCCCGCT GCAGGAGGTG GAGATGGAGA CCTATGGGGC 5940

CGGCACCCCC TTGGAGTCCT GTGCCATCCC ATCCAGATCC CATTGGCTGT GTCGAACCCA 6000

GCCAGGAGCG GCGAGCCCCT CCACGCCCTG TCCCCTCGGG GCACAGCCGC TCCCCCAGTC 6060

TCAGCCGGCT GCTCTGCAGA CAGGAGGCTG TGCACACCGA TTCCTTGGAA GGGAAGATTG 6120

ACAGCCCTAG GGACACCCTG GATCCTGCAG AGCCTGGTGA GAAACCCCCG GTGAGGCCGG 6180

TGACCCAGGG GGGCTCCCTG CAGTCCCCAC CACGCTCCCC ACGGCCCGCC AGCGTCCGCA 6240

CTCGTAAGCA TACCTTCGGA CAGCGCTGCG TCTCCAGCCG GCCGGCGGCC CCAGGCGGAG 6300

AGGAGGCCGA GGCCTCGGAC CCAGCCGACG AGGAGGTCAG CCACATCACC AGCTCCGCCT 6360

GCCCCTGGCA GCCCACAGCC GAGCCCCATG GCCCCGAAGC CTCTCCGGTG GCCGGCGGCG 6420

AGCGGGACCT GCGCAGGCTC TACAGCGTGG ATGCTCAGGG CTTCCTGGAC AAGCCGGGCC 6480

GGGCAGACGA GCAGTGGCTG CCCTCGGGGA GTGGGCAGCG GGGAGCCTGG GGAGGCGAAG 6540

GCCTGGGGCC TGAGGCCGAG CCCGCTCTGG GTGCGCGCAG AAAGAAGAAG ATGAGCCCCC 6600

CCTGCATCTC GGTGGAACCC CCTGCGGAGG ACGAGGGCTC TGCGCGGCCC TCCGCGGCAG 6660

AGGGCGGCAG ACCACACTGA GGCTCAGGAC CCCGTCCTGT GAGGCCACGC CTCACAGGGA 6720

CTCCCTGGAG CCCACAGAGG GCTCAGGCGC CGGGGGGGAC CCTGCAGCCA AGGGGGAGCG 6780

CTGGGGCCAG GCCTCCTGCC GGGCTGAGCA CCTGACCGTC CCCAGCTTTG CCTTTGAGCC 6840

GCTGGACCTC GGGGTCCCCA GTGGAGACCC TTTCTTGGAC GGTAGCCACA GTGTGACCCC 6900

AGAATCCAGA GCTTCCTCTT CAGGGGCCAT AGTGCCCCTG GAACCCCCAG AATCAGAGCC 6960

TCCCATGCCC GTCGGTGACC CCCCAGAGAA GAGGCGGGGG CTGTACCTCA CAGTCCCCCA 7020

GTGTCCTCTG GAGAAACCAG GGTCCCCCTC AGCCACCCCT GCCCCAGGGG GTGGTGCAGA 7080

TGACCCCGTG TAGCTCGGGG CTTGGTGCCG CCCACGGCTT TGGCCCTGGG GTCTGGGGGC 7140

CCGCTGGGGT GGAGGCCCAG GCAGAACCCT GCATGGACCC TGACTTGGGT CCCGTCGTGA 7200

GCAGAAAGGC CCGGGGAGGA TGACGGCCCA GGCCCTGGTT CTCTGCCCAG CGAAGCAGGA 7260

GTAGCTGCCG GGCCCCCACG AGCCTCCGTC CGTTCTGGTT CGGGTTTCTC CGAGTTTTGC 7320

TACCAGCCGA GGCTGTGCCG GCAACTGGGT CAGCCTCCCG TCAGGAGAGA AGCCGCGTCT 7380

GTGGGACGAA GACCGGGCAC CCGCCAGAGA GGGGAATGGT ACCAGGTTGC GTCCTTTCAG 7440

GCCCCGCGTT GTTACAGGAT CATCTCGCTG GGGGCCCTGT GCCTCTTGCC GGCGGCAGGT 7500

TGCATGCCAC CGCGGCCCGA ATGTCACCTT CACTCACAGT CTGAGTTCTT GTCCGCCTGT 7560

CACGCCCTCA CCACCCTCCC CTTCCAGCCA CCACCCTTTC CGTTCCGCTC GGGCCTTCCC 7620

AGAAGCGTCC TGTGACTCTG GGAGAGGTGA CACCTCACTA AGGGGCCGAC CCCATGGAGT 7680

AACGCGCCCG GCCCCGATGC GAATCAGGCC TCCCCCTCCG 7720

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6858 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATG CTC CCC CAC CGG GTC CCC CGT TGC GTG AGG ACA CCT CCT CTG AGG 48
Met Leu Pro His Arg Val Pro Arg Cys Val Arg Thr Pro Pro Leu Arg
2035 2040 2045

GGC TCC GCT CGC CCC TCT TCG GAC CCC CCG GGG CCC CGG CTG GCC AGA 96
Gly Ser Ala Arg Pro Ser Ser Asp Pro Pro Gly Pro Arg Leu Ala Arg
2050 2055 2060

GGA TGG ACG AGG AGG AGG ATG GAG CGG GCG CCG AGG AGT CGG GAC AGC 144
Gly Trp Thr Arg Arg Arg Met Glu Arg Ala Pro Arg Ser Arg Asp Ser
2065 2070 2075 2080

CCC GTA GCT TCA CGC AGC TCA ACG ACC TGT CCG GGG CCG GGG GCG GCA 192

```
Pro Val Ala Ser Arg Ser Ser Thr Thr Cys Pro Gly Pro Gly Ala Ala
                2085                2090                2095

GGG GCC GGG TCG ACG GAA AAG GAC CCG GGC AGC GCG GAC TCC GAG GCG      240
Gly Ala Gly Ser Thr Glu Lys Asp Pro Gly Ser Ala Asp Ser Glu Ala
            2100                2105                2110

GAG GGG CTG CCG TAC CCG GCG CTA GCC CCG GTG GTT TTC TTC TAC TTG      288
Glu Gly Leu Pro Tyr Pro Ala Leu Ala Pro Val Val Phe Phe Tyr Leu
        2115                2120                2125

AGC CAG GAC AGC CGC CCG CGG AGC TGG TGT CTC CGC ACG GTC TGT AAC      336
Ser Gln Asp Ser Arg Pro Arg Ser Trp Cys Leu Arg Thr Val Cys Asn
    2130                2135                2140

CCG TGG TTC GAG CGA GTC AGT ATG CTG GTC ATT CTT CTC AAC TGT GTG      384
Pro Trp Phe Glu Arg Val Ser Met Leu Val Ile Leu Leu Asn Cys Val
2145                2150                2155                2160

ACT CTG GGT ATG TTC AGG CCG TGT GAG GAC ATT GCC TGT GAC TCC CAG      432
Thr Leu Gly Met Phe Arg Pro Cys Glu Asp Ile Ala Cys Asp Ser Gln
                2165                2170                2175

CGC TGC CGG ATC CTG CAG GCC TTC GAT GAC TTC ATC TTT GCC TTC TTT      480
Arg Cys Arg Ile Leu Gln Ala Phe Asp Asp Phe Ile Phe Ala Phe Phe
            2180                2185                2190

GCT GTG GAA ATG GTG GTG AAG ATG GTG GCC TTG GGC ATC TTT GGG AAG      528
Ala Val Glu Met Val Val Lys Met Val Ala Leu Gly Ile Phe Gly Lys
        2195                2200                2205

AAA TGT TAC CTG GGA GAC ACT TGG AAC CGG CTT GAC TTT TTC ATT GTC      576
Lys Cys Tyr Leu Gly Asp Thr Trp Asn Arg Leu Asp Phe Phe Ile Val
    2210                2215                2220

ATT GCA GGG ATG CTG GAG TAT TCG CTG GAC CTG CAG AAC GTC AGC TTC      624
Ile Ala Gly Met Leu Glu Tyr Ser Leu Asp Leu Gln Asn Val Ser Phe
```

2225 2230 2235 2240

TCC GCA GTC AGG ACA GTC CGT GTG CTG CGA CCG CTC AGG GCC ATT AAC 672
Ser Ala Val Arg Thr Val Arg Val Leu Arg Pro Leu Arg Ala Ile Asn
2245 2250 2255

CGG GTG CCC AGC ATG CGC ATT CTC GTC ACA TTA CTG CTG GAC ACC TTG 720
Arg Val Pro Ser Met Arg Ile Leu Val Thr Leu Leu Leu Asp Thr Leu
2260 2265 2270

CCT ATG CTG GGC AAC GTC CTG CTG CTC TGT TTC TTC GTC TTT TTC ATC 768
Pro Met Leu Gly Asn Val Leu Leu Leu Cys Phe Phe Val Phe Phe Ile
2275 2280 2285

TTT GGC ATC GTG GGC GTC CAG CTG TGG GCA GGA CTG CTT CGC AAC CGG 816
Phe Gly Ile Val Gly Val Gln Leu Trp Ala Gly Leu Leu Arg Asn Arg
2290 2295 2300

TGC TTC CTC CCC GAG AAC TTC AGC CTC CCC CTG AGC GTG GAC CTG GAG 864
Cys Phe Leu Pro Glu Asn Phe Ser Leu Pro Leu Ser Val Asp Leu Glu
2305 2310 2315 2320

CCT TAT TAC CAG ACA GAG AAT GAG GAC GAG AGC CCC TTC ATC TGC TCT 912
Pro Tyr Tyr Gln Thr Glu Asn Glu Asp Glu Ser Pro Phe Ile Cys Ser
2325 2330 2335

CAG CCT CGG GAG AAT GGC ATG AGA TCC TGC AGG AGT GTG CCC ACA CTG 960
Gln Pro Arg Glu Asn Gly Met Arg Ser Cys Arg Ser Val Pro Thr Leu
2340 2345 2350

CGT GGG GAA GGC GGT GGT GGC CCA CCC TGC AGT CTG GAC TAT GAG ACC 1008
Arg Gly Glu Gly Gly Gly Gly Pro Pro Cys Ser Leu Asp Tyr Glu Thr
2355 2360 2365

TAT AAC AGT TCC AGC AAC ACC ACC TGT GTC AAC TGG AAC CAG TAC TAT 1056
Tyr Asn Ser Ser Ser Asn Thr Thr Cys Val Asn Trp Asn Gln Tyr Tyr
2370 2375 2380

```
ACC AAC TGC TCT GCG GGC GAG CAC AAC CCC TTC AAA GGC GCC ATC AAC      1104
Thr Asn Cys Ser Ala Gly Glu His Asn Pro Phe Lys Gly Ala Ile Asn
2385                2390                2395                2400

TTT GAC AAC ATT GGC TAT GCC TGG ATC GCC ATC TTC CAG GTC ATC ACA      1152
Phe Asp Asn Ile Gly Tyr Ala Trp Ile Ala Ile Phe Gln Val Ile Thr
                2405                2410                2415

CTG GAG GGC TGG GTC GAC ATC ATG TAC TTC GTA ATG GAC GCT CAC TCC      1200
Leu Glu Gly Trp Val Asp Ile Met Tyr Phe Val Met Asp Ala His Ser
            2420                2425                2430

TTC TAC AAC TTC ATC TAC TTC ATT CTT CTC ATC ATC GTG GGC TCC TTC      1248
Phe Tyr Asn Phe Ile Tyr Phe Ile Leu Leu Ile Ile Val Gly Ser Phe
        2435                2440                2445

TTC ATG ATC AAC CTG TGC CTG GTG GTG ATT GCC ACG CAG TTC TCC GAG      1296
Phe Met Ile Asn Leu Cys Leu Val Val Ile Ala Thr Gln Phe Ser Glu
    2450                2455                2460

ACC AAA CAG CGG GAG AGT CAG CTG ATG CGG GAG CAG CGT GTA CGA TTC      1344
Thr Lys Gln Arg Glu Ser Gln Leu Met Arg Glu Gln Arg Val Arg Phe
2465                2470                2475                2480

CTG TCC AAT GCT AGC ACC CTG GCA AGC TTC TCT GAG CCA GGC AGC TGC      1392
Leu Ser Asn Ala Ser Thr Leu Ala Ser Phe Ser Glu Pro Gly Ser Cys
                2485                2490                2495

TAT GAG GAG CTA CTC AAG TAC CTG GTG TAC ATC CTC CGA AAA GCA GCC      1440
Tyr Glu Glu Leu Leu Lys Tyr Leu Val Tyr Ile Leu Arg Lys Ala Ala
            2500                2505                2510

CGA AGG CTG GCC CAG GTC TCT AGG GCT ATA GGC GTG CGG GCT GGG CTG      1488
Arg Arg Leu Ala Gln Val Ser Arg Ala Ile Gly Val Arg Ala Gly Leu
        2515                2520                2525
```

```
CTC AGC AGC CCA GTG GCC CGT AGT GGG CAG GAG CCC CAG CCC AGT GGC      1536
Leu Ser Ser Pro Val Ala Arg Ser Gly Gln Glu Pro Gln Pro Ser Gly
    2530                2535                2540

AGC TGC ACT CGC TCA CAC CGT CGT CTG TCT GTC CAC CAC CTG GTC CAC      1584
Ser Cys Thr Arg Ser His Arg Arg Leu Ser Val His His Leu Val His
2545                2550                2555                2560

CAC CAT CAC CAC CAC CAT CAC CAC TAC CAC CTG GGT AAT GGG ACG CTC      1632
His His His His His His His His Tyr His Leu Gly Asn Gly Thr Leu
                2565                2570                2575

AGA GTT CCC CGG GCC AGC CCA GAG ATC CAG GAC AGG GAT GCC AAT GGG      1680
Arg Val Pro Arg Ala Ser Pro Glu Ile Gln Asp Arg Asp Ala Asn Gly
            2580                2585                2590

TCT CGC CGG CTC ATG CTA CCA CCA CCC TCT ACA CCC ACT CCC TCT GGG      1728
Ser Arg Arg Leu Met Leu Pro Pro Pro Ser Thr Pro Thr Pro Ser Gly
        2595                2600                2605

GGC CCT CCG AGG GGT GCG GAG TCT GTA CAC AGC TTC TAC CAT GCT GAC      1776
Gly Pro Pro Arg Gly Ala Glu Ser Val His Ser Phe Tyr His Ala Asp
    2610                2615                2620

TGC CAC TTG GAG CCA GTC CGT TGC CAG GCA CCC CCT CCC AGA TGC CCA      1824
Cys His Leu Glu Pro Val Arg Cys Gln Ala Pro Pro Pro Arg Cys Pro
2625                2630                2635                2640

TCG GAG GCA TCT GGT AGG ACT GTG GGT AGT GGG AAG GTG TAC CCC ACT      1872
Ser Glu Ala Ser Gly Arg Thr Val Gly Ser Gly Lys Val Tyr Pro Thr
                2645                2650                2655

GTG CAT ACC AGC CCT CCA CCA GAG ATA CTG AAG GAT AAA GCA CTA GTG      1920
Val His Thr Ser Pro Pro Pro Glu Ile Leu Lys Asp Lys Ala Leu Val
            2660                2665                2670

GAG GTG GCC CCC AGC CCT GGG CCC CCC ACC CTC ACC AGC TTC AAC ATC      1968
```

```
Glu Val Ala Pro Ser Pro Gly Pro Pro Thr Leu Thr Ser Phe Asn Ile
        2675                2680                2685

CCA CCT GGG CCC TTC AGC TCC ATG CAC AAG CTC CTG GAG ACA CAG AGT      2016
Pro Pro Gly Pro Phe Ser Ser Met His Lys Leu Leu Glu Thr Gln Ser
    2690                2695                2700

ACG GGA GCC TGC CAT AGC TCC TGC AAA ATC TCC AGC CCT TGC TCC AAG      2064
Thr Gly Ala Cys His Ser Ser Cys Lys Ile Ser Ser Pro Cys Ser Lys
2705                2710                2715                2720

GCA GAC AGT GGA GCC TGC GGG CCG GAC AGT TGT CCC TAC TGT GCC CGG      2112
Ala Asp Ser Gly Ala Cys Gly Pro Asp Ser Cys Pro Tyr Cys Ala Arg
                2725                2730                2735

ACA GGA GCA GGA GAG CCA GAG TCC GCT GAC CAT GTC ATG CCT GAC TCA      2160
Thr Gly Ala Gly Glu Pro Glu Ser Ala Asp His Val Met Pro Asp Ser
            2740                2745                2750

GAC AGC GAG GCT GTG TAT GAG TTC ACA CAG GAC GCT CAG CAC AGT GAC      2208
Asp Ser Glu Ala Val Tyr Glu Phe Thr Gln Asp Ala Gln His Ser Asp
        2755                2760                2765

CTC CGG GAT CCC CAC AGC CGG CGG CGA CAG CGG AGC CTG GGC CCA GAT      2256
Leu Arg Asp Pro His Ser Arg Arg Arg Gln Arg Ser Leu Gly Pro Asp
    2770                2775                2780

GCA GAG CCT AGT TCT GTG CTG GCT TTC TGG AGG CTG ATC TGT GAC ACA      2304
Ala Glu Pro Ser Ser Val Leu Ala Phe Trp Arg Leu Ile Cys Asp Thr
2785                2790                2795                2800

TTC CGG AAG ATC GTA GAT AGC AAA TAC TTT GGC CGG GGA ATC ATG ATC      2352
Phe Arg Lys Ile Val Asp Ser Lys Tyr Phe Gly Arg Gly Ile Met Ile
                2805                2810                2815

GCC ATC CTG GTC AAT ACA CTC AGC ATG GGC ATC GAG TAC CAC GAG CAG      2400
Ala Ile Leu Val Asn Thr Leu Ser Met Gly Ile Glu Tyr His Glu Gln
```

```
                2820                2825                2830

CCC GAG GAG CTC ACC AAC GCC CTG GAA ATC AGC AAC ATC GTC TTC ACC      2448
Pro Glu Glu Leu Thr Asn Ala Leu Glu Ile Ser Asn Ile Val Phe Thr
        2835                2840                2845

AGC CTC TTC GCC TTG GAG ATG CTG CTG AAA CTG CTT GTC TAC GGT CCC      2496
Ser Leu Phe Ala Leu Glu Met Leu Leu Lys Leu Leu Val Tyr Gly Pro
    2850                2855                2860

TTT GGC TAC ATT AAG AAT CCC TAC AAC ATC TTT GAT GGT GTC ATT GTG      2544
Phe Gly Tyr Ile Lys Asn Pro Tyr Asn Ile Phe Asp Gly Val Ile Val
2865                2870                2875                2880

GTC ATC AGT GTG TGG GAG ATT GTG GGC CAG CAG GGA GGT GGC CTG TCG      2592
Val Ile Ser Val Trp Glu Ile Val Gly Gln Gln Gly Gly Gly Leu Ser
                2885                2890                2895

GTG CTG CGG ACC TTC CGC CTG ATG CGG GTG CTG AAG CTG GTG CGC TTC      2640
Val Leu Arg Thr Phe Arg Leu Met Arg Val Leu Lys Leu Val Arg Phe
            2900                2905                2910

CTG CCG GCC CTG CAG CGC CAG CTC GTG GTG CTC ATG AAG ACC ATG GAC      2688
Leu Pro Ala Leu Gln Arg Gln Leu Val Val Leu Met Lys Thr Met Asp
        2915                2920                2925

AAC GTG GCC ACC TTC TGC ATG CTC CTC ATG CTG TTC ATC TTC ATC TTC      2736
Asn Val Ala Thr Phe Cys Met Leu Leu Met Leu Phe Ile Phe Ile Phe
    2930                2935                2940

AGC ATC CTG GGC ATG CAT CTC TTT GGT TGC AAG TTC GCA TCT GAA CGG      2784
Ser Ile Leu Gly Met His Leu Phe Gly Cys Lys Phe Ala Ser Glu Arg
2945                2950                2955                2960

GAT GGG GAC ACG TTG CCA GAC CGG AAG AAT TTC GAC TCC CTG CTC TGG      2832
Asp Gly Asp Thr Leu Pro Asp Arg Lys Asn Phe Asp Ser Leu Leu Trp
                2965                2970                2975
```

```
GCC ATC GTC ACT GTC TTT CAG ATT CTG ACT CAG GAA GAC TGG AAT AAA      2880
Ala Ile Val Thr Val Phe Gln Ile Leu Thr Gln Glu Asp Trp Asn Lys
            2980                2985                2990

GTC CTC TAC AAC GGC ATG GCC TCC ACA TCG TCT TGG GCT GCT CTT TAC      2928
Val Leu Tyr Asn Gly Met Ala Ser Thr Ser Ser Trp Ala Ala Leu Tyr
        2995                3000                3005

TTC ATC GCC CTC ATG ACT TTT GGC AAC TAT GTG CTC TTT AAC CTG CTG      2976
Phe Ile Ala Leu Met Thr Phe Gly Asn Tyr Val Leu Phe Asn Leu Leu
    3010                3015                3020

GTG GCC ATT CTT GTG GAA GGA TTC CAG GCA GAG GGA GAT GCC ACC AAG      3024
Val Ala Ile Leu Val Glu Gly Phe Gln Ala Glu Gly Asp Ala Thr Lys
3025                3030                3035                3040

TCT GAG TCA GAG CCT GAT TTC TTT TCG CCC AGT GTG GAT GGT GAT GGG      3072
Ser Glu Ser Glu Pro Asp Phe Phe Ser Pro Ser Val Asp Gly Asp Gly
                3045                3050                3055

GAC AGA AAG AAG CGC TTG GCC CTG GTG GCT TTG GGA GAA CAC GCG GAA      3120
Asp Arg Lys Lys Arg Leu Ala Leu Val Ala Leu Gly Glu His Ala Glu
            3060                3065                3070

CTA CGA AAG AGC CTT TTG CCA CCC CTC ATC ATC CAT ACG GCT GCG ACA      3168
Leu Arg Lys Ser Leu Leu Pro Pro Leu Ile Ile His Thr Ala Ala Thr
        3075                3080                3085

CCA ATG TCA CAC CCC AAG AGC TCC AGC ACA GGT GTG GGG GAA GCA CTG      3216
Pro Met Ser His Pro Lys Ser Ser Ser Thr Gly Val Gly Glu Ala Leu
    3090                3095                3100

GGC TCT GGC TCT CGA CGT ACC AGT AGC AGT GGG TCC GCT GAG CCT GGA      3264
Gly Ser Gly Ser Arg Arg Thr Ser Ser Ser Gly Ser Ala Glu Pro Gly
3105                3110                3115                3120
```

```
GCT GCC CAC CAT GAG ATG AAA TGT CCG CCA AGT GCC CGC AGC TCC CCG      3312
Ala Ala His His Glu Met Lys Cys Pro Pro Ser Ala Arg Ser Ser Pro
                3125                3130                3135

CAC AGT CCC TGG AGT GCG GCA AGC AGC TGG ACC AGC AGG CGC TCC AGC      3360
His Ser Pro Trp Ser Ala Ala Ser Ser Trp Thr Ser Arg Arg Ser Ser
            3140                3145                3150

AGG AAC AGC CTG GGC CGG GCC CCC AGC CTA AAG CGG AGG AGC CCG AGC      3408
Arg Asn Ser Leu Gly Arg Ala Pro Ser Leu Lys Arg Arg Ser Pro Ser
        3155                3160                3165

GGG GAG CGG AGG TCC CTG CTG TCT GGA GAG GGC CAG GAG AGT CAG GAT      3456
Gly Glu Arg Arg Ser Leu Leu Ser Gly Glu Gly Gln Glu Ser Gln Asp
    3170                3175                3180

GAG GAG GAA AGT TCA GAA GAG GAC CGG GCC AGC CCA GCA GGC AGT GAC      3504
Glu Glu Glu Ser Ser Glu Glu Asp Arg Ala Ser Pro Ala Gly Ser Asp
3185                3190                3195                3200

CAT CGC CAC AGG GGT TCC TTG GAA CGT GAG GCC AAG AGT TCC TTT GAC      3552
His Arg His Arg Gly Ser Leu Glu Arg Glu Ala Lys Ser Ser Phe Asp
                3205                3210                3215

CTG CCT GAC ACT CTG CAG GTG CCG GGG CTG CAC CGC ACA GCC AGC GGC      3600
Leu Pro Asp Thr Leu Gln Val Pro Gly Leu His Arg Thr Ala Ser Gly
            3220                3225                3230

CGG AGC TCT GCC TCT GAG CAC CAA GAC TGT AAT GGC AAG TCG GCT TCA      3648
Arg Ser Ser Ala Ser Glu His Gln Asp Cys Asn Gly Lys Ser Ala Ser
        3235                3240                3245

GGG CGT TTG GCC CGC ACC CTG AGG ACT GAT GAC CCC CAA CTG GAT GGG      3696
Gly Arg Leu Ala Arg Thr Leu Arg Thr Asp Asp Pro Gln Leu Asp Gly
    3250                3255                3260

GAT GAT GAC AAT GAT GAG GGA AAT CTG AGC AAA GGG GAA CGC ATA CAA      3744
```

```
Asp Asp Asp Asn Asp Glu Gly Asn Leu Ser Lys Gly Glu Arg Ile Gln
3265                3270                3275                3280

GCC TGG GTC AGA TCC CGG CTT CCT GCC TGT TGC CGA GAG CGA GAT TCC      3792
Ala Trp Val Arg Ser Arg Leu Pro Ala Cys Cys Arg Glu Arg Asp Ser
                3285                3290                3295

TGG TCG GCC TAT ATC TTT CCT CCT CAG TCA AGG TTT CGT CTC CTG TGT      3840
Trp Ser Ala Tyr Ile Phe Pro Pro Gln Ser Arg Phe Arg Leu Leu Cys
            3300                3305                3310

CAC CGG ATC ATC ACC CAC AAG ATG TTT GAC CAT GTG GTC CTC GTC ATC      3888
His Arg Ile Ile Thr His Lys Met Phe Asp His Val Val Leu Val Ile
        3315                3320                3325

ATC TTC CTC AAC TGT ATC ACC ATC GCT ATG GAG CGC CCC AAA ATT GAC      3936
Ile Phe Leu Asn Cys Ile Thr Ile Ala Met Glu Arg Pro Lys Ile Asp
    3330                3335                3340

CCC CAC AGC GCT GAG CGC ATC TTC CTG ACC CTC TCC AAC TAC ATC TTC      3984
Pro His Ser Ala Glu Arg Ile Phe Leu Thr Leu Ser Asn Tyr Ile Phe
3345                3350                3355                3360

ACG GCA GTC TTT CTA GCT GAA ATG ACA GTG AAG GTG GTG GCA CTG GGC      4032
Thr Ala Val Phe Leu Ala Glu Met Thr Val Lys Val Val Ala Leu Gly
                3365                3370                3375

TGG TGC TTT GGG GAG CAG GCC TAC CTG CGC AGC AGC TGG AAT GTG CTG      4080
Trp Cys Phe Gly Glu Gln Ala Tyr Leu Arg Ser Ser Trp Asn Val Leu
            3380                3385                3390

GAC GGC TTG CTG GTG CTC ATC TCC GTC ATC GAC ATC CTG GTC TCC ATG      4128
Asp Gly Leu Leu Val Leu Ile Ser Val Ile Asp Ile Leu Val Ser Met
        3395                3400                3405

GTC TCC GAC AGC GGC ACC AAG ATC CTT GGC ATG CTG AGG GTG CTG CGG      4176
Val Ser Asp Ser Gly Thr Lys Ile Leu Gly Met Leu Arg Val Leu Arg
```

```
    3410                3415                3420

CTG CTG CGG ACC CTG CGC CCA CTC AGG GTC ATC AGC CGG GCC CAG GGA      4224
Leu Leu Arg Thr Leu Arg Pro Leu Arg Val Ile Ser Arg Ala Gln Gly
3425                3430                3435                3440

CTG AAG CTG GTG GTA GAG ACT CTG ATG TCA TCC CTC AAA CCC ATT GGC      4272
Leu Lys Leu Val Val Glu Thr Leu Met Ser Ser Leu Lys Pro Ile Gly
                3445                3450                3455

AAC ATT GTG GTC ATT TGC TGT GCC TTC TTC ATC ATT TTT GGA ATT CTC      4320
Asn Ile Val Val Ile Cys Cys Ala Phe Phe Ile Ile Phe Gly Ile Leu
            3460                3465                3470

GGG GTG CAG CTC TTC AAA GGG AAG TTC TTC GTG TGT CAG GGT GAG GAC      4368
Gly Val Gln Leu Phe Lys Gly Lys Phe Phe Val Cys Gln Gly Glu Asp
        3475                3480                3485

ACC AGG AAC ATC ACT AAC AAA TCC GAC TGC GCT GAG GCC AGC TAC CGA      4416
Thr Arg Asn Ile Thr Asn Lys Ser Asp Cys Ala Glu Ala Ser Tyr Arg
    3490                3495                3500

TGG GTC CGG CAC AAG TAC AAC TTT GAC AAC CTG GGC CAG GCT CTG ATG      4464
Trp Val Arg His Lys Tyr Asn Phe Asp Asn Leu Gly Gln Ala Leu Met
3505                3510                3515                3520

TCC CTG TTT GTG CTG GCC TCC AAG GAT GGT TGG GTT GAC ATC ATG TAT      4512
Ser Leu Phe Val Leu Ala Ser Lys Asp Gly Trp Val Asp Ile Met Tyr
                3525                3530                3535

GAT GGG CTG GAT GCT GTG GGT GTG GAT CAG CAG CCC ATC ATG AAC CAC      4560
Asp Gly Leu Asp Ala Val Gly Val Asp Gln Gln Pro Ile Met Asn His
            3540                3545                3550

AAC CCC TGG ATG CTG CTA TAC TTC ATC TCC TTC CTC CTC ATC GTG GCC      4608
Asn Pro Trp Met Leu Leu Tyr Phe Ile Ser Phe Leu Leu Ile Val Ala
        3555                3560                3565
```

```
TTC TTT GTC CTG AAC ATG TTT GTG GGC GTG GTG GTG GAG AAC TTC CAT      4656
Phe Phe Val Leu Asn Met Phe Val Gly Val Val Val Glu Asn Phe His
    3570                3575                3580

AAG TGC AGA CAG CAC CAG GAG GAG GAG GAG GCG AGG CGG CGT GAG GAG      4704
Lys Cys Arg Gln His Gln Glu Glu Glu Glu Ala Arg Arg Arg Glu Glu
3585                3590                3595                3600

AAG CGA CTA CGG AGG CTG GAG AAA AAG AGA AGG AGT AAG GAG AAG CAG      4752
Lys Arg Leu Arg Arg Leu Glu Lys Lys Arg Arg Ser Lys Glu Lys Gln
                3605                3610                3615

ATG GCC GAA GCC CAG TGC AAG CCC TAC TAC TCT GAC TAC TCG AGA TTC      4800
Met Ala Glu Ala Gln Cys Lys Pro Tyr Tyr Ser Asp Tyr Ser Arg Phe
            3620                3625                3630

CGG CTC CTT GTC CAC CAC CTG TGT ACC AGC CAC TAC CTG GAC CTC TTC      4848
Arg Leu Leu Val His His Leu Cys Thr Ser His Tyr Leu Asp Leu Phe
        3635                3640                3645

ATC ACT GGT GTC ATC GGG CTG AAC GTG GTC ACT ATG GCC ATG GAA CAT      4896
Ile Thr Gly Val Ile Gly Leu Asn Val Val Thr Met Ala Met Glu His
    3650                3655                3660

TAC CAG CAG CCC CAG ATC CTG GAC GAG GCT CTG AAG ATC TGC AAT TAC      4944
Tyr Gln Gln Pro Gln Ile Leu Asp Glu Ala Leu Lys Ile Cys Asn Tyr
3665                3670                3675                3680

ATC TTT ACC GTC ATC TTT GTC TTT GAG TCA GTT TTC AAA CTT GTG GCC      4992
Ile Phe Thr Val Ile Phe Val Phe Glu Ser Val Phe Lys Leu Val Ala
                3685                3690                3695

TTT GCG TTC CGC CGT TTC TTC CAG GAC AGG TGG AAC CAG CTG GAC CTG      5040
Phe Ala Phe Arg Arg Phe Phe Gln Asp Arg Trp Asn Gln Leu Asp Leu
            3700                3705                3710
```

```
GCT ATT GTG CTT CTG TCC ATC ATG GGC ATC ACA CTG GAG GAG ATT GAG      5088
Ala Ile Val Leu Leu Ser Ile Met Gly Ile Thr Leu Glu Glu Ile Glu
        3715                3720                3725

GTC AAT CTG TCG CTG CCC ATC AAC CCC ACC ATC ATC CGT ATC ATG AGG      5136
Val Asn Leu Ser Leu Pro Ile Asn Pro Thr Ile Ile Arg Ile Met Arg
    3730                3735                3740

GTG CTC CGC ATT GCT CGA GTT CTG AAG CTG TTG AAG ATG GCT GTG GGC      5184
Val Leu Arg Ile Ala Arg Val Leu Lys Leu Leu Lys Met Ala Val Gly
3745                3750                3755                3760

ATG CGG GCA CTG CTG CAC ACG GTG ATG CAG GCC CTG CCC CAG GTG GGG      5232
Met Arg Ala Leu Leu His Thr Val Met Gln Ala Leu Pro Gln Val Gly
                3765                3770                3775

AAC CTG GGA CTT CTC TTC ATG TTA TTG TTT TTC ATC TTT GCA GCT CTG      5280
Asn Leu Gly Leu Leu Phe Met Leu Leu Phe Phe Ile Phe Ala Ala Leu
            3780                3785                3790

GGC GTG GAG CTC TTT GGA GAC CTG GAG TGT GAT GAG ACA CAC CCT TGT      5328
Gly Val Glu Leu Phe Gly Asp Leu Glu Cys Asp Glu Thr His Pro Cys
        3795                3800                3805

GAG GGC TTG GGT CGG CAT GCC ACC TTT AGG AAC TTT GGT ATG GCC TTT      5376
Glu Gly Leu Gly Arg His Ala Thr Phe Arg Asn Phe Gly Met Ala Phe
    3810                3815                3820

CTG ACC CTC TTC CGA GTC TCC ACT GGT GAC AAC TGG AAT GGT ATT ATG      5424
Leu Thr Leu Phe Arg Val Ser Thr Gly Asp Asn Trp Asn Gly Ile Met
3825                3830                3835                3840

AAG GAC CCT TCC CGG GAC TGT GAC CAG GAG TCC ACC TGC TAC AAC ACT      5472
Lys Asp Pro Ser Arg Asp Cys Asp Gln Glu Ser Thr Cys Tyr Asn Thr
                3845                3850                3855

GTC ATC TCC CCT ATC TAC TTT GTG TCC TTC GTG CTG ACG GCC CAG TTT      5520
```

```
Val Ile Ser Pro Ile Tyr Phe Val Ser Phe Val Leu Thr Ala Gln Phe
            3860                3865                3870

GTG CTG GTC AAC GTG GTC ATA GCT GTG CTG ATG AAG CAC CTG GAA GAA      5568
Val Leu Val Asn Val Val Ile Ala Val Leu Met Lys His Leu Glu Glu
        3875                3880                3885

AGC AAC AAA GAG GCC AAG GAG GAG GCC GAG CTC GAG GCC GAG CTG GAG      5616
Ser Asn Lys Glu Ala Lys Glu Glu Ala Glu Leu Glu Ala Glu Leu Glu
    3890                3895                3900

CTG GAG ATG AAG ACG CTC AGC CCG CAG CCC CAC TCC CCG CTG GGC AGC      5664
Leu Glu Met Lys Thr Leu Ser Pro Gln Pro His Ser Pro Leu Gly Ser
3905                3910                3915                3920

CCC TTC CTC TGG CCC GGG GTG GAG GGT GTC AAC AGT ACT GAC AGC CCT      5712
Pro Phe Leu Trp Pro Gly Val Glu Gly Val Asn Ser Thr Asp Ser Pro
                3925                3930                3935

AAG CCT GGG GCT CCA CAC ACC ACT GCC CAC ATT GGA GCA GCC TCG GGC      5760
Lys Pro Gly Ala Pro His Thr Thr Ala His Ile Gly Ala Ala Ser Gly
            3940                3945                3950

TTC TCC CTT GAG CAC CCC ACG ATG GTA CCC CAC CCC GAG GAG GTG CCA      5808
Phe Ser Leu Glu His Pro Thr Met Val Pro His Pro Glu Glu Val Pro
        3955                3960                3965

GTC CCC CTA GGA CCA GAC CTG CTG ACT GTG AGG AAG TCT GGT GTC AGC      5856
Val Pro Leu Gly Pro Asp Leu Leu Thr Val Arg Lys Ser Gly Val Ser
    3970                3975                3980

CGG ACG CAC TCT CTG CCC AAT GAC AGC TAC ATG TGC CGC AAT GGG AGC      5904
Arg Thr His Ser Leu Pro Asn Asp Ser Tyr Met Cys Arg Asn Gly Ser
3985                3990                3995                4000

ACT GCT GAG AGA TCC CTA GGA CAC AGG GGC TGG GGG CTC CCC AAA GCC      5952
Thr Ala Glu Arg Ser Leu Gly His Arg Gly Trp Gly Leu Pro Lys Ala
```

4005 4010 4015

CAG TCA GGC TCC ATC TTG TCC GTT CAC TCC CAA CCA GCA GAC ACC AGC 6000
Gln Ser Gly Ser Ile Leu Ser Val His Ser Gln Pro Ala Asp Thr Ser
4020 4025 4030

TGC ATC CTA CAG CTT CCC AAA GAT GTG CAC TAT CTG CTC CAG CCT CAT 6048
Cys Ile Leu Gln Leu Pro Lys Asp Val His Tyr Leu Leu Gln Pro His
4035 4040 4045

GGG GCT CCC ACC TGG GGC GCC ATC CCT AAA CTA CCC CCA CCT GGC CGC 6096
Gly Ala Pro Thr Trp Gly Ala Ile Pro Lys Leu Pro Pro Pro Gly Arg
4050 4055 4060

TCC CCT CTG GCT CAG AGG CCT CTC AGG CGC CAG GCA GCA ATA AGG ACT 6144
Ser Pro Leu Ala Gln Arg Pro Leu Arg Arg Gln Ala Ala Ile Arg Thr
4065 4070 4075 4080

GAC TCC CTG GAT GTG CAG GGC CTG GGT AGC CGG GAA GAC CTG TTG TCA 6192
Asp Ser Leu Asp Val Gln Gly Leu Gly Ser Arg Glu Asp Leu Leu Ser
4085 4090 4095

GAG GTG AGT GGG CCC TCC TGC CCT CTG ACC CGG TCC TCA TCC TTC TGG 6240
Glu Val Ser Gly Pro Ser Cys Pro Leu Thr Arg Ser Ser Ser Phe Trp
4100 4105 4110

GGC GGG TCG AGC ATC CAG GTG CAG CAG CGT TCC GGC ATC CAG AGC AAA 6288
Gly Gly Ser Ser Ile Gln Val Gln Gln Arg Ser Gly Ile Gln Ser Lys
4115 4120 4125

GTC TCC AAG CAC ATC CGC CTG CCA GCC CCT TGC CCA GGC CTG GAA CCC 6336
Val Ser Lys His Ile Arg Leu Pro Ala Pro Cys Pro Gly Leu Glu Pro
4130 4135 4140

AGC TGG GCC AAG GAC CCT CCA GAG ACC AGA AGC AGC TTA GAG CTG GAC 6384
Ser Trp Ala Lys Asp Pro Pro Glu Thr Arg Ser Ser Leu Glu Leu Asp
4145 4150 4155 4160

```
ACG GAG CTG AGC TGG ATT TCA GGA GAC CTC CTT CCC AGC AGC CAG GAA      6432
Thr Glu Leu Ser Trp Ile Ser Gly Asp Leu Leu Pro Ser Ser Gln Glu
                4165                4170                4175

GAA CCC CTG TTC CCA CGG GAC CTG AAG AAG TGC TAC AGT GTA GAG ACC      6480
Glu Pro Leu Phe Pro Arg Asp Leu Lys Lys Cys Tyr Ser Val Glu Thr
            4180                4185                4190

CAG AGC TGC AGG CGC AGG CCT GGG TTC TGG CTA GAT GAA CAG CGG AGA      6528
Gln Ser Cys Arg Arg Arg Pro Gly Phe Trp Leu Asp Glu Gln Arg Arg
        4195                4200                4205

CAC TCC ATT GCT GTC AGC TGT CTG GAC AGC GGC TCC CAA CCC CGC CTA      6576
His Ser Ile Ala Val Ser Cys Leu Asp Ser Gly Ser Gln Pro Arg Leu
    4210                4215                4220

TGT CCA AGC CCC TCA AGC CTC GGG GGC CAA CCT CTT GGG GGT CCT GGG      6624
Cys Pro Ser Pro Ser Ser Leu Gly Gly Gln Pro Leu Gly Gly Pro Gly
4225                4230                4235                4240

AGC CGG CCT AAG AAA AAA CTC AGC CCA CCC AGT ATC TCT ATA GAC CCC      6672
Ser Arg Pro Lys Lys Lys Leu Ser Pro Pro Ser Ile Ser Ile Asp Pro
                4245                4250                4255

CCG GAG AGC CAG GGC TCT CGG CCC CCA TGC AGT CCT GGT GTC TGC CTC      6720
Pro Glu Ser Gln Gly Ser Arg Pro Pro Cys Ser Pro Gly Val Cys Leu
            4260                4265                4270

AGG AGG AGG GCG CCG GCC AGT GAC TCT AAG GAT CCC TCG GTC TCC AGC      6768
Arg Arg Arg Ala Pro Ala Ser Asp Ser Lys Asp Pro Ser Val Ser Ser
        4275                4280                4285

CCC CTT GAC AGC ACG GCT GCC TCA CCC TCC CCA AAG AAA GAC ACG CTG      6816
Pro Leu Asp Ser Thr Ala Ala Ser Pro Ser Pro Lys Lys Asp Thr Leu
    4290                4295                4300
```

AGT CTC TCT GGT TTG TCT TCT GAC CCA ACA GAC ATG GAC CCC 6858
Ser Leu Ser Gly Leu Ser Ser Asp Pro Thr Asp Met Asp Pro
4305 4310 4315

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CCGTCTCTGG CGCGGAGCGG GACGATGCTG ACCCCTTAGA TCCTGCTCCA GCTGCGCCGA 60

GGGAAGAGGG GGCGCCCCTC CCCGGACCCC CGCCCTCCAT CGGGTGGCCC CTTTTTTTTC 120

TCTTCCTCTC GGGGGCTGCC TCGCCGAAGG TAGCGCCTGT TACGGGCAAC CGGAGCCTGG 180

GCGCGAACGA AGAAGCCGGA ACAAAGTGAG GGGAAGCCGC CCGGCTAGTC GGGGAGCCCC 240

CGGGAACCCA GGGGAAGCGG GACTCTACGC CAGGCGGGGC TTCCCTGAGA CCCGGCGCCC 300

CGCGGGCAGC ATGCCCTGAG GGCAGGGGGA GCTGAGCTGA ACTGGCCCTC CTGGGGACTC 360

AGCAAGCTCT CTAGAGCCCC CCACATGCTC CCCCACCGGG TCCCCCGTTG CGTGAGGACA 420

CCTCCTCTGA GGGGCTCCGC TCGCCCCTCT TCGGACCCCC CGGGGCCCCG GCTGGCCAGA 480

GGATGGACGA GGAGGAGGAT GGAGCGGGCG CCGAGGAGTC GGGACAGCCC CGTAGCTTCA 540

CGCAGCTCAA CGACCTGTCC GGGGCCGGGG GCGGCAGGGG CCGGGTCGAC GGAAAAGGAC 600

CCGGGCAGCG CGGACTCCGA GGCGGAGGGG CTGCCGTACC CGGCGCTAGC CCCGGTGGTT 660

TTCTTCTACT TGAGCCAGGA CAGCCGCCCG CGGAGCTGGT GTCTCCGCAC GGTCTGTAAC 720

CCGTGGTTCG AGCGAGTCAG TATGCTGGTC ATTCTTCTCA ACTGTGTGAC TCTGGGTATG 780

TTCAGGCCGT GTGAGGACAT TGCCTGTGAC TCCCAGCGCT GCCGGATCCT GCAGGCCTTC 840

GATGACTTCA TCTTTGCCTT CTTTGCTGTG GAAATGGTGG TGAAGATGGT GGCCTTGGGC 900

ATCTTTGGGA AGAAATGTTA CCTGGGAGAC ACTTGGAACC GGCTTGACTT TTTCATTGTC 960

ATTGCAGGGA TGCTGGAGTA TTCGCTGGAC CTGCAGAACG TCAGCTTCTC CGCAGTCAGG 1020

ACAGTCCGTG TGCTGCGACC GCTCAGGGCC ATTAACCGGG TGCCCAGCAT GCGCATTCTC 1080

GTCACATTAC TGCTGGACAC CTTGCCTATG CTGGGCAACG TCCTGCTGCT CTGTTTCTTC 1140

GTCTTTTTCA TCTTTGGCAT CGTGGGCGTC CAGCTGTGGG CAGGACTGCT TCGCAACCGG 1200

TGCTTCCTCC CCGAGAACTT CAGCCTCCCC CTGAGCGTGG ACCTGGAGCC TTATTACCAG 1260

ACAGAGAATG AGGACGAGAG CCCCTTCATC TGCTCTCAGC CTCGGGAGAA TGGCATGAGA 1320

TCCTGCAGGA GTGTGCCCAC ACTGCGTGGG GAAGGCGGTG GTGGCCCACC CTGCAGTCTG 1380

GACTATGAGA CCTATAACAG TTCCAGCAAC ACCACCTGTG TCAACTGGAA CCAGTACTAT 1440

ACCAACTGCT CTGCGGGCGA GCACAACCCC TTCAAAGGCG CCATCAACTT TGACAACATT 1500

GGCTATGCCT GGATCGCCAT CTTCCAGGTC ATCACACTGG AGGGCTGGGT CGACATCATG 1560

TACTTCGTAA TGGACGCTCA CTCCTTCTAC AACTTCATCT ACTTCATTCT TCTCATCATC 1620

GTGGGCTCCT TCTTCATGAT CAACCTGTGC CTGGTGGTGA TTGCCACGCA GTTCTCCGAG 1680

```
ACCAAACAGC GGGAGAGTCA GCTGATGCGG GAGCAGCGTG TACGATTCCT GTCCAATGCT    1740

AGCACCCTGG CAAGCTTCTC TGAGCCAGGC AGCTGCTATG AGGAGCTACT CAAGTACCTG    1800

GTGTACATCC TCCGAAAAGC AGCCCGAAGG CTGGCCCAGG TCTCTAGGGC TATAGGCGTG    1860

CGGGCTGGGC TGCTCAGCAG CCCAGTGGCC CGTAGTGGGC AGGAGCCCCA GCCCAGTGGC    1920

AGCTGCACTC GCTCACACCG TCGTCTGTCT GTCCACCACC TGGTCCACCA CCATCACCAC    1980

CACCATCACC ACTACCACCT GGGTAATGGG ACGCTCAGAG TTCCCCGGGC CAGCCCAGAG    2040

ATCCAGGACA GGGATGCCAA TGGGTCTCGC CGGCTCATGC TACCACCACC CTCTACACCC    2100

ACTCCCTCTG GGGGCCCTCC GAGGGGTGCG GAGTCTGTAC ACAGCTTCTA CCATGCTGAC    2160

TGCCACTTGG AGCCAGTCCG TTGCCAGGCA CCCCCTCCCA GATGCCCATC GGAGGCATCT    2220

GGTAGGACTG TGGGTAGTGG GAAGGTGTAC CCCACTGTGC ATACCAGCCC TCCACCAGAG    2280

ATACTGAAGG ATAAAGCACT AGTGGAGGTG GCCCCCAGCC CTGGGCCCCC CACCCTCACC    2340

AGCTTCAACA TCCCACCTGG GCCCTTCAGC TCCATGCACA AGCTCCTGGA GACACAGAGT    2400

ACGGGAGCCT GCCATAGCTC CTGCAAAATC TCCAGCCCTT GCTCCAAGGC AGACAGTGGA    2460

GCCTGCGGGC CGGACAGTTG TCCCTACTGT GCCCGGACAG GAGCAGGAGA GCCAGAGTCC    2520

GCCGACCATG TCATGCCTGA CTCAGACAGC GAGGCCGTGT ATGAGTTCAC ACAGGACGCT    2580

CAGCACAGTG ACCTCCGGGA TCCCCACAGC CGGCGGCGAC AGCGGAGCCT GGGCCCAGAT    2640

GCAGAGCCTA GTTCTGTGCT GGCTTTCTGG AGGCTGATCT GTGACACATT CCGGAAGATC    2700

GTAGATAGCA AATACTTTGG CCGGGGAATC ATGATCGCCA TCCTGGTCAA TACACTCAGC    2760

ATGGGCATCG AGTACCACGA GCAGCCCGAG GAGCTCACCA ACGCCCTGGA AATCAGCAAC    2820
```

ATCGTCTTCA CCAGCCTCTT CGCCTTGGAG ATGCTGCTGA AACTGCTTGT CTACGGTCCC 2880

TTTGGCTACA TTAAGAATCC CTACAACATC TTTGATGGTG TCATTGTGGT CATCAGTGTG 2940

TGGGAGATTG TGGGCCAGCA GGGAGGTGGC CTGTCGGTGC TGCGGACCTT CCGCCTGATG 3000

CGGGTGCTGA AGCTGGTGCG CTTCCTGCCG GCCCTGCAGC GCCAGCTCGT GGTGCTCATG 3060

AAGACCATGG ACAACGTGGC CACCTTCTGC ATGCTCCTCA TGCTGTTCAT CTTCATCTTC 3120

AGCATCCTGG GCATGCATCT CTTTGGTTGC AAGTTCGCAT CTGAACGGGA TGGGGACACG 3180

TTGCCAGACC GGAAGAATTT CGACTCCCTG CTCTGGGCCA TCGTCACTGT CTTTCAGATT 3240

CTGACTCAGG AAGACTGGAA TAAAGTCCTC TACAACGGCA TGGCCTCCAC ATCGTCTTGG 3300

GCTGCTCTTT ACTTCATCGC CCTCATGACT TTTGGCAACT ATGTGCTCTT TAACCTGCTG 3360

GTGGCCATTC TTGTGGAAGG ATTCCAGGCA GAGGGAGATG CCACCAAGTC TGAGTCAGAG 3420

CCTGATTTCT TTTCGCCCAG TGTGGATGGT GATGGGGACA GAAAGAAGCG CTTGGCCCTG 3480

GTGGCTTTGG GAGAACACGC GGAACTACGA AAGAGCCTTT TGCCACCCCT CATCATCCAT 3540

ACGGCTGCGA CACCAATGTC ACACCCCAAG AGCTCCAGCA CAGGTGTGGG GGAAGCACTG 3600

GGCTCTGGCT CTCGACGTAC CAGTAGCAGT GGGTCCGCTG AGCCTGGAGC TGCCCACCAT 3660

GAGATGAAAT GTCCGCCAAG TGCCCGCAGC TCCCCGCACA GTCCCTGGAG TGCGGCAAGC 3720

AGCTGGACCA GCAGGCGCTC CAGCAGGAAC AGCCTGGGCC GGGCCCCCAG CCTAAAGCGG 3780

AGGAGCCCGA GCGGGGAGCG GAGGTCCCTG CTGTCTGGAG AGGGCCAGGA GAGTCAGGAT 3840

GAGGAGGAAA GTTCAGAAGA GGACCGGGCC AGCCCAGCAG GCAGTGACCA TCGCCACAGG 3900

GGTTCCTTGG AACGTGAGGC CAAGAGTTCC TTTGACCTGC CTGACACTCT GCAGGTGCCG 3960

GGGCTGCACC GCACAGCCAG CGGCCGGAGC TCTGCCTCTG AGCACCAAGA CTGTAATGGC 4020

AAGTCGGCTT CAGGGCGTTC GGCCCGCACC CTGAGGACTG ATGACCCCCA ACTGGATGGG 4080

GATGATGACA ATGATGAGGG AAATCTGAGC AAAGGGGAAC GCATACAAGC CTGGGTCAGA 4140

TCCCGGCTTC CTGCCTGTTG CCGAGAGCGA GATTCCTGGT CGGCCTATAT CTTTCCTCCT 4200

CAGTCAAGGT TTCGTCTCCT GTGTCACCGG ATCATCACCC ACAAGATGTT TGACCATGTG 4260

GTCCTCGTCA TCATCTTCCT CAACTGTACC ACCATCGCTA TGGAGCGCCC CAAAATTGAC 4320

CCCCACAGCG CTGAGCGCAT CTTCCTGACC CTCTCCAACT ACATCTTCAC GGCAGTCTTT 4380

CTAGCTGAAA TGACAGTGAA GGTGGTGGCA CTGGGCTGGT GCTTTGGGGA GCAGGCCTAC 4440

CTGCGCAGCA GCTGGAATGT GCTGGACGGC TTGCTGGTGC TCATCTCCGT CATCGACATC 4500

CTGGTCTCCA TGGTCTCCGA CAGCGGCACC AAGATCCTTG GCATGCTGAG GGTGCTGCGG 4560

CTGCTGCGGA CCCTGCGTCC ACTCAGGGTC ATCAGCCGGG CCCAGGGACT GAAGCTGGTG 4620

GTAGAGACTC TGATGTCATC CCTCAAACCC ATTGGCAACA TTGTGGTCAT TTGCTGTGCC 4680

TTCTTCATCA TTTTTGGAAT TCTCGGGGTG CAGCTCTTCA AAGGGAAGTT CTTCGTGTGT 4740

CAGGGTGAGG ACACCAGGAA CATCACTAAC AAATCCGACT GCGCTGAGGC CAGCTACCGA 4800

TGGGTCCGGC ACAAGTACAA CTTTGACAAC CTGGGCCAGG CTCTGATGTC CCTGTTTGTG 4860

CTGGCCTCCA AGGATGGTTG GGTTGACATC ATGTATGATG GGCTGGATGC TGTGGGTGTG 4920

GATCAGCAGC CCATCATGAA CCACAACCCC TGGATGCTGC TATACTTCAT CTCCTTCCTC 4980

CTCATCGTGG CCTTCTTTGT CCTGAACATG TTTGTGGGCG TGGTGGTGGA GAACTTCCAT 5040

AAGTGCAGAC AGCACCAGGA GGAGGAGGAG GCGAGGCGGC GTGAGGAGAA GCGACTACGG 5100

AGGCTGGAGA AAAAGAGAAG GAGTAAGGAG AAGCAGATGG CCGAAGCCCA GTGCAAGCCC 5160

TACTACTCTG ACTACTCGAG ATTCCGGCTC CTTGTCCACC ACCTGTGTAC CAGCCACTAC 5220

CTGGACCTCT TCATCACTGG TGTCATCGGG CTGAACGTGG TCACTATGGC CATGGAACAT 5280

TACCAGCAGC CCCAGATCCT GGACGAGGCT CTGAAGATCT GCAATTACAT CTTTACCGTC 5340

ATCTTTGTCT TTGAGTCAGT TTTCAAACTT GTGGCCTTTG CGTTCCGCCG TTTCTTCCAG 5400

GACAGGTGGA ACCAGCTGGA CCTGGCTATT GTGCTTCTGT CCATCATGGG CATCACACTG 5460

GAGGAGATTG AGGTCAATCT GTCGCTGCCC ATCAACCCCA CCATCATCCG TATCATGAGG 5520

GTGCTCCGCA TTGCTCGAGT TCTGAAGCTG TTGAAGATGG CTGTGGGCAT GCGGGCACTG 5580

CTGCACACGG TGATGCAGGC CCTGCCCCAG GTGGGGAACC TGGGACTTCT CTTCATGTTA 5640

TTGTTTTTCA TCTTTGCAGC TCTGGGCGTG GAGCTCTTTG GAGACCTGGA GTGTGATGAG 5700

ACACACCCTT GTGAGGGCTT GGGTCGGCAT GCCACCTTTA GGAACTTTGG TATGGCCTTT 5760

CTGACCCTCT TCCGAGTCTC CACTGGTGAC AACTGGAATG GTATTATGAA GGACCCTTCC 5820

CGGGACTGTG ACCAGGAGTC CACCTGCTAC AACACTGTCA TCTCCCCTAT CTACTTTGTG 5880

TCCTTCGTGC TGACGGCCCA GTTTGTGCTG GTCAACGTGG TCATAGCTGT GCTGATGAAG 5940

CACCTGGAAG AAAGCAACAA AGAGGCCAAG GAGGAGGCCG AGCTCGAGGC CGAGCTGGAG 6000

CTGGAGATGA AGACGCTCAG CCCGCAGCCC CACTCCCCGC TGGGCAGCCC CTTCCTCTGG 6060

CCCGGGGTGG AGGGTGTCAA CAGTACTGAC AGCCCTAAGC CTGGGGCTCC ACACACCACT 6120

GCCCACATTG GAGCAGCCTC GGGCTTCTCC CTTGAGCACC CCACGATGGT ACCCCACCCC 6180

GAGGAGGTGC CAGTCCCCCT AGGACCAGAC CTGCTGACTG TGAGGAAGTC TGGTGTCAGC 6240

CGGACGCACT CTCTGCCCAA TGACAGCTAC ATGTGCCGCA ATGGGAGCAC TGCTGAGAGA 6300

TCCCTAGGAC ACAGGGGCTG GGGGCTCCCC AAAGCCCAGT CAGGCTCCAT CTTGTCCGTT 6360

CACTCCCAAC CAGCAGACAC CAGCTGCATC CTACAGCTTC CCAAAGATGT GCACTATCTG 6420

CTCCAGCCTC ATGGGGCTCC CACCTGGGGC GCCATCCCTA AACTACCCCC ACCTGGCCGC 6480

TCCCCTCTGG CTCAGAGGCC TCTCAGGCGC CAGCCAGCAA TAAGGACTGA CTCCCTGGAT 6540

GTGCAGGGCC TGGGTAGCCG GGAAGACCTG TTGTCAGAGG TGAGTGGGCC CTCCTGCCCT 6600

CTGACCCGGT CCTCATCCTT CTGGGGCGGG TCGAGCATCC AGGTGCAGCA GCGTTCCGGC 6660

ATCCAGAGCA AAGTCTCCAA GCACATCCGC CTGCCAGCCC CTTGCCCAGG CCTGGAACCC 6720

AGCTGGGCCA AGGACCCTCC AGAGACCAGA AGCAGCTTAG AGCTGGACAC GGAGCTGAGC 6780

TGGATTTCAG GAGACCTCCT TCCCAGCAGC CAGGAAGAAC CCCTGTTCCC ACGGGACCTG 6840

AAGAAGTGCT ACAGTGTAGA GACCCAGAGC TGCAGGCGCA GGCCTGGGTT CTGGCTAGAT 6900

GAACAGCGGA GACACTCCAT TGCTGTCAGC TGTCTGGACA GCGGCTCCCA ACCCCGCCTA 6960

TGTCCAAGCC CCTCAAGCCT CGGGGGCCAA CCTCTTGGGG GTCCTGGGAG CCGGCCTAAG 7020

AAAAAACTCA GCCCACCCAG TATCTCTATA GACCCCCCGG AGAGCCAGGG CTCTCGGCCC 7080

CCATGCAGTC CTGGTGTCTG CCTCAGGAGG AGGGCGCCGG CCAGTGACTC TAAGGATCCC 7140

TCGGTCTCCA GCCCCCTTGA CAGCACGGCT GCCTCACCCT CCCCAAAGAA AGACACGCTG 7200

AGTCTCTCTG GTTTGTCTTC TGACCCAACA GACATGGACC CCTGAGTCCT ACCCACTCTC 7260

CCCCATCACC TTTCTCCACC GGGTGCAGAT CCTACGTCCG CCTCCTGGGC AGCGTTTCTG 7320

AAAAGTCCCA CGTAAGCAGC AAGCAGCCAC GAGGCACCTC ACCTGCCTTC TTCAGTGGCT 7380

GGTGGGGATG ACGAGCAGAA CTTCCGGAGA GTCGATCTGA AGAGAACACA GCCCTGGAGC 7440

CCCTGCCTCC GGGAAGAAGG AAAAGGAGAA GCCCAGTGTG GCCAAGGCTC CCGACACCAG 7500

GAGCTGTTGG GAGAAGCAAT ACGTTTGTGC AGAATCTCTA 7540

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2297 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Leu Pro His Arg Val Pro Arg Cys Val Arg Thr Pro Pro Leu Arg
1 5 10 15

Gly Ser Ala Arg Pro Ser Ser Asp Pro Pro Gly Pro Arg Leu Ala Arg
20 25 30

Gly Trp Thr Arg Arg Arg Met Glu Arg Ala Pro Arg Ser Arg Asp Ser
35 40 45

Pro Val Ala Ser Arg Ser Ser Thr Thr Cys Pro Gly Pro Gly Ala Ala

65

50 55 60

Gly Ala Gly Ser Thr Glu Lys Asp Pro Gly Ser Ala Asp Ser Glu Ala
65 70 75 80

Glu Gly Leu Pro Tyr Pro Ala Leu Ala Pro Val Val Phe Phe Tyr Leu
85 90 95

Ser Gln Asp Ser Arg Pro Arg Ser Trp Cys Leu Arg Thr Val Cys Asn
100 105 110

Pro Trp Phe Glu Arg Val Ser Met Leu Val Ile Leu Leu Asn Cys Val
115 120 125

Thr Leu Gly Met Phe Arg Pro Cys Glu Asp Ile Ala Cys Asp Ser Gln
130 135 140

Arg Cys Arg Ile Leu Gln Ala Phe Asp Asp Phe Ile Phe Ala Phe Phe
145 150 155 160

Ala Val Glu Met Val Val Lys Met Val Ala Leu Gly Ile Phe Gly Lys
165 170 175

Lys Cys Tyr Leu Gly Asp Thr Trp Asn Arg Leu Asp Phe Phe Ile Val
180 185 190

Ile Ala Gly Met Leu Glu Tyr Ser Leu Asp Leu Gln Asn Val Ser Phe
195 200 205

Ser Ala Val Arg Thr Val Arg Val Leu Arg Pro Leu Arg Ala Ile Asn
210 215 220

Arg Val Pro Ser Met Arg Ile Leu Val Thr Leu Leu Leu Asp Thr Leu
225 230 235 240

Pro Met Leu Gly Asn Val Leu Leu Leu Cys Phe Phe Val Phe Phe Ile
245 250 255

```
Phe Gly Ile Val Gly Val Gln Leu Trp Ala Gly Leu Leu Arg Asn Arg
            260                 265                 270

Cys Phe Leu Pro Glu Asn Phe Ser Leu Pro Leu Ser Val Asp Leu Glu
        275                 280                 285

Pro Tyr Tyr Gln Thr Glu Asn Glu Asp Glu Ser Pro Phe Ile Cys Ser
    290                 295                 300

Gln Pro Arg Glu Asn Gly Met Arg Ser Cys Arg Ser Val Pro Thr Leu
305                 310                 315                 320

Arg Gly Glu Gly Gly Gly Gly Pro Pro Cys Ser Leu Asp Tyr Glu Thr
                325                 330                 335

Tyr Asn Ser Ser Ser Asn Thr Thr Cys Val Asn Trp Asn Gln Tyr Tyr
            340                 345                 350

Thr Asn Cys Ser Ala Gly Glu His Asn Pro Phe Lys Gly Ala Ile Asn
        355                 360                 365

Phe Asp Asn Ile Gly Tyr Ala Trp Ile Ala Ile Phe Gln Val Ile Thr
    370                 375                 380

Leu Glu Gly Trp Val Asp Ile Met Tyr Phe Val Met Asp Ala His Ser
385                 390                 395                 400

Phe Tyr Asn Phe Ile Tyr Phe Ile Leu Leu Ile Ile Val Gly Ser Phe
                405                 410                 415

Phe Met Ile Asn Leu Cys Leu Val Val Ile Ala Thr Gln Phe Ser Glu
            420                 425                 430

Thr Lys Gln Arg Glu Ser Gln Leu Met Arg Glu Gln Arg Val Arg Phe
        435                 440                 445
```

Leu Ser Asn Ala Ser Thr Leu Ala Ser Phe Ser Glu Pro Gly Ser Cys
450 455 460

Tyr Glu Glu Leu Leu Lys Tyr Leu Val Tyr Ile Leu Arg Lys Ala Ala
465 470 475 480

Arg Arg Leu Ala Gln Val Ser Arg Ala Ile Gly Val Arg Ala Gly Leu
485 490 495

Leu Ser Ser Pro Val Ala Arg Ser Gly Gln Glu Pro Gln Pro Ser Gly
500 505 510

Ser Cys Thr Arg Ser His Arg Arg Leu Ser Val His His Leu Val His
515 520 525

His His His His His His His His Tyr His Leu Gly Asn Gly Thr Leu
530 535 540

Arg Val Pro Arg Ala Ser Pro Glu Ile Gln Asp Arg Asp Ala Asn Gly
545 550 555 560

Ser Arg Arg Leu Met Leu Pro Pro Pro Ser Thr Pro Thr Pro Ser Gly
565 570 575

Gly Pro Pro Arg Gly Ala Glu Ser Val His Ser Phe Tyr His Ala Asp
580 585 590

Cys His Leu Glu Pro Val Arg Cys Gln Ala Pro Pro Pro Arg Cys Pro
595 600 605

Ser Glu Ala Ser Gly Arg Thr Val Gly Ser Gly Lys Val Tyr Pro Thr
610 615 620

Val His Thr Ser Pro Pro Pro Glu Ile Leu Lys Asp Lys Ala Leu Val
625 630 635 640

Glu Val Ala Pro Ser Pro Gly Pro Pro Thr Leu Thr Ser Phe Asn Ile

645 650 655

Pro Pro Gly Pro Phe Ser Ser Met His Lys Leu Leu Glu Thr Gln Ser
660 665 670

Thr Gly Ala Cys His Ser Ser Cys Lys Ile Ser Ser Pro Cys Ser Lys
675 680 685

Ala Asp Ser Gly Ala Cys Gly Pro Asp Ser Cys Pro Tyr Cys Ala Arg
690 695 700

Thr Gly Ala Gly Glu Pro Glu Ser Ala Asp His Val Met Pro Asp Ser
705 710 715 720

Asp Ser Glu Ala Val Tyr Glu Phe Thr Gln Asp Ala Gln His Ser Asp
725 730 735

Leu Arg Asp Pro His Ser Arg Arg Arg Gln Arg Ser Leu Gly Pro Asp
740 745 750

Ala Glu Pro Ser Ser Val Leu Ala Phe Trp Arg Leu Ile Cys Asp Thr
755 760 765

Phe Arg Lys Ile Val Asp Ser Lys Tyr Phe Gly Arg Gly Ile Met Ile
770 775 780

Ala Ile Leu Val Asn Thr Leu Ser Met Gly Ile Glu Tyr His Glu Gln
785 790 795 800

Pro Glu Glu Leu Thr Asn Ala Leu Glu Ile Ser Asn Ile Val Phe Thr
805 810 815

Ser Leu Phe Ala Leu Glu Met Leu Leu Lys Leu Leu Val Tyr Gly Pro
820 825 830

Phe Gly Tyr Ile Lys Asn Pro Tyr Asn Ile Phe Asp Gly Val Ile Val
835 840 845

Val Ile Ser Val Trp Glu Ile Val Gly Gln Gln Gly Gly Gly Leu Ser
850 855 860

Val Leu Arg Thr Phe Arg Leu Met Arg Val Leu Lys Leu Val Arg Phe
865 870 875 880

Leu Pro Ala Leu Gln Arg Gln Leu Val Val Leu Met Lys Thr Met Asp
885 890 895

Asn Val Ala Thr Phe Cys Met Leu Leu Met Leu Phe Ile Phe Ile Phe
900 905 910

Ser Ile Leu Gly Met His Leu Phe Gly Cys Lys Phe Ala Ser Glu Arg
915 920 925

Asp Gly Asp Thr Leu Pro Asp Arg Lys Asn Phe Asp Ser Leu Leu Trp
930 935 940

Ala Ile Val Thr Val Phe Gln Ile Leu Thr Gln Glu Asp Trp Asn Lys
945 950 955 960

Val Leu Tyr Asn Gly Met Ala Ser Thr Ser Ser Trp Ala Ala Leu Tyr
965 970 975

Phe Ile Ala Leu Met Thr Phe Gly Asn Tyr Val Leu Phe Asn Leu Leu
980 985 990

Val Ala Ile Leu Val Glu Gly Phe Gln Ala Glu Gly Asp Ala Thr Lys
995 1000 1005

Ser Glu Ser Glu Pro Asp Phe Phe Ser Pro Ser Val Asp Gly Asp Gly
1010 1015 1020

Asp Arg Lys Lys Arg Leu Ala Leu Val Ala Leu Gly Glu His Ala Glu
1025 1030 1035 1040

Leu Arg Lys Ser Leu Leu Pro Pro Leu Ile Ile His Thr Ala Ala Thr
1045 1050 1055

Pro Met Ser His Pro Lys Ser Ser Ser Thr Gly Val Gly Glu Ala Leu
1060 1065 1070

Gly Ser Gly Ser Arg Arg Thr Ser Ser Ser Gly Ser Ala Glu Pro Gly
1075 1080 1085

Ala Ala His His Glu Met Lys Cys Pro Pro Ser Ala Arg Ser Ser Pro
1090 1095 1100

His Ser Pro Trp Ser Ala Ala Ser Ser Trp Thr Ser Arg Arg Ser Ser
1105 1110 1115 1120

Arg Asn Ser Leu Gly Arg Ala Pro Ser Leu Lys Arg Arg Ser Pro Ser
1125 1130 1135

Gly Glu Arg Arg Ser Leu Leu Ser Gly Glu Gly Gln Glu Ser Gln Asp
1140 1145 1150

Glu Glu Glu Ser Ser Glu Glu Asp Arg Ala Ser Pro Ala Gly Ser Asp
1155 1160 1165

His Arg His Arg Gly Ser Leu Glu Arg Glu Ala Lys Ser Ser Phe Asp
1170 1175 1180

Leu Pro Asp Thr Leu Gln Val Pro Gly Leu His Arg Thr Ala Ser Gly
1185 1190 1195 1200

Arg Ser Ser Ala Ser Glu His Gln Asp Cys Asn Gly Lys Ser Ala Ser
1205 1210 1215

Gly Arg Leu Ala Arg Thr Leu Arg Thr Asp Asp Pro Gln Leu Asp Gly
1220 1225 1230

Asp Asp Asp Asn Asp Glu Gly Asn Leu Ser Lys Gly Glu Arg Ile Gln

1235 1240 1245

Ala Trp Val Arg Ser Arg Leu Pro Ala Cys Cys Arg Glu Arg Asp Ser
1250 1255 1260

Trp Ser Ala Tyr Ile Phe Pro Pro Gln Ser Arg Phe Arg Leu Leu Cys
1265 1270 1275 1280

His Arg Ile Ile Thr His Lys Met Phe Asp His Val Val Leu Val Ile
1285 1290 1295

Ile Phe Leu Asn Cys Ile Thr Ile Ala Met Glu Arg Pro Lys Ile Asp
1300 1305 1310

Pro His Ser Ala Glu Arg Ile Phe Leu Thr Leu Ser Asn Tyr Ile Phe
1315 1320 1325

Thr Ala Val Phe Leu Ala Glu Met Thr Val Lys Val Val Ala Leu Gly
1330 1335 1340

Trp Cys Phe Gly Glu Gln Ala Tyr Leu Arg Ser Ser Trp Asn Val Leu
1345 1350 1355 1360

Asp Gly Leu Leu Val Leu Ile Ser Val Ile Asp Ile Leu Val Ser Met
1365 1370 1375

Val Ser Asp Ser Gly Thr Lys Ile Leu Gly Met Leu Arg Val Leu Arg
1380 1385 1390

Leu Leu Arg Thr Leu Arg Pro Leu Arg Val Ile Ser Arg Ala Gln Gly
1395 1400 1405

Leu Lys Leu Val Val Glu Thr Leu Met Ser Ser Leu Lys Pro Ile Gly
1410 1415 1420

Asn Ile Val Val Ile Cys Cys Ala Phe Phe Ile Ile Phe Gly Ile Leu
1425 1430 1435 1440

Gly Val Gln Leu Phe Lys Gly Lys Phe Phe Val Cys Gln Gly Glu Asp
1445 1450 1455

Thr Arg Asn Ile Thr Asn Lys Ser Asp Cys Ala Glu Ala Ser Tyr Arg
1460 1465 1470

Trp Val Arg His Lys Tyr Asn Phe Asp Asn Leu Gly Gln Ala Leu Met
1475 1480 1485

Ser Leu Phe Val Leu Ala Ser Lys Asp Gly Trp Val Asp Ile Met Tyr
1490 1495 1500

Asp Gly Leu Asp Ala Val Gly Val Asp Gln Gln Pro Ile Met Asn His
1505 1510 1515 1520

Asn Pro Trp Met Leu Leu Tyr Phe Ile Ser Phe Leu Leu Ile Val Ala
1525 1530 1535

Phe Phe Val Leu Asn Met Phe Val Gly Val Val Val Glu Asn Phe His
1540 1545 1550

Lys Cys Arg Gln His Gln Glu Glu Glu Glu Ala Arg Arg Arg Glu Glu
1555 1560 1565

Lys Arg Leu Arg Arg Leu Glu Lys Lys Arg Arg Asn Leu Met Leu Asp
1570 1575 1580

Asp Val Ile Ala Ser Gly Ser Ser Ala Ser Ala Ala Ser Glu Ala Gln
1585 1590 1595 1600

Cys Lys Pro Tyr Tyr Ser Asp Tyr Ser Arg Phe Arg Leu Leu Val His
1605 1610 1615

His Leu Cys Thr Ser His Tyr Leu Asp Leu Phe Ile Thr Gly Val Ile
1620 1625 1630

Gly Leu Asn Val Val Thr Met Ala Met Glu His Tyr Gln Gln Pro Gln
1635 1640 1645

Ile Leu Asp Glu Ala Leu Lys Ile Cys Asn Tyr Ile Phe Thr Val Ile
1650 1655 1660

Phe Val Phe Glu Ser Val Phe Lys Leu Val Ala Phe Ala Phe Arg Arg
1665 1670 1675 1680

Phe Phe Gln Asp Arg Trp Asn Gln Leu Asp Leu Ala Ile Val Leu Leu
1685 1690 1695

Ser Ile Met Gly Ile Thr Leu Glu Glu Ile Glu Val Asn Leu Ser Leu
1700 1705 1710

Pro Ile Asn Pro Thr Ile Ile Arg Ile Met Arg Val Leu Arg Ile Ala
1715 1720 1725

Arg Val Leu Lys Leu Leu Lys Met Ala Val Gly Met Arg Ala Leu Leu
1730 1735 1740

His Thr Val Met Gln Ala Leu Pro Gln Val Gly Asn Leu Gly Leu Leu
1745 1750 1755 1760

Phe Met Leu Leu Phe Phe Ile Phe Ala Ala Leu Gly Val Glu Leu Phe
1765 1770 1775

Gly Asp Leu Glu Cys Asp Glu Thr His Pro Cys Glu Gly Leu Gly Arg
1780 1785 1790

His Ala Thr Phe Arg Asn Phe Gly Met Ala Phe Leu Thr Leu Phe Arg
1795 1800 1805

Val Ser Thr Gly Asp Asn Trp Asn Gly Ile Met Lys Asp Pro Ser Arg
1810 1815 1820

Asp Cys Asp Gln Glu Ser Thr Cys Tyr Asn Thr Val Ile Ser Pro Ile

1825 1830 1835 1840

Tyr Phe Val Ser Phe Val Leu Thr Ala Gln Phe Val Leu Val Asn Val
1845 1850 1855

Val Ile Ala Val Leu Met Lys His Leu Glu Glu Ser Asn Lys Glu Ala
1860 1865 1870

Lys Glu Glu Ala Glu Leu Glu Ala Glu Leu Glu Leu Glu Met Lys Thr
1875 1880 1885

Leu Ser Pro Gln Pro His Ser Pro Leu Gly Ser Pro Phe Leu Trp Pro
1890 1895 1900

Gly Val Glu Gly Val Asn Ser Thr Asp Ser Pro Lys Pro Gly Ala Pro
1905 1910 1915 1920

His Thr Thr Ala His Ile Gly Ala Ala Ser Gly Phe Ser Leu Glu His
1925 1930 1935

Pro Thr Met Val Pro His Pro Glu Glu Val Pro Val Pro Leu Gly Pro
1940 1945 1950

Asp Leu Leu Thr Val Arg Lys Ser Gly Val Ser Arg Thr His Ser Leu
1955 1960 1965

Pro Asn Asp Ser Tyr Met Cys Arg Asn Gly Ser Thr Ala Glu Arg Ser
1970 1975 1980

Leu Gly His Arg Gly Trp Gly Leu Pro Lys Ala Gln Ser Gly Ser Ile
1985 1990 1995 2000

Leu Ser Val His Ser Gln Pro Ala Asp Thr Ser Cys Ile Leu Gln Leu
2005 2010 2015

Pro Lys Asp Val His Tyr Leu Leu Gln Pro His Gly Ala Pro Thr Trp
2020 2025 2030

Gly Ala Ile Pro Lys Leu Pro Pro Pro Gly Arg Ser Pro Leu Ala Gln
2035 2040 2045

Arg Pro Leu Arg Arg Gln Ala Ala Ile Arg Thr Asp Ser Leu Asp Val
2050 2055 2060

Gln Gly Leu Gly Ser Arg Glu Asp Leu Leu Ser Glu Val Ser Gly Pro
2065 2070 2075 2080

Ser Cys Pro Leu Thr Arg Ser Ser Ser Phe Trp Gly Gly Ser Ser Ile
2085 2090 2095

Gln Val Gln Gln Arg Ser Gly Ile Gln Ser Lys Val Ser Lys His Ile
2100 2105 2110

Arg Leu Pro Ala Pro Cys Pro Gly Leu Glu Pro Ser Trp Ala Lys Asp
2115 2120 2125

Pro Pro Glu Thr Arg Ser Ser Leu Glu Leu Asp Thr Glu Leu Ser Trp
2130 2135 2140

Ile Ser Gly Asp Leu Leu Pro Ser Ser Gln Glu Glu Pro Leu Phe Pro
2145 2150 2155 2160

Arg Asp Leu Lys Lys Cys Tyr Ser Val Glu Thr Gln Ser Cys Arg Arg
2165 2170 2175

Arg Pro Gly Phe Trp Leu Asp Glu Gln Arg Arg His Ser Ile Ala Val
2180 2185 2190

Ser Cys Leu Asp Ser Gly Ser Gln Pro Arg Leu Cys Pro Ser Pro Ser
2195 2200 2205

Ser Leu Gly Gly Gln Pro Leu Gly Gly Pro Gly Ser Arg Pro Lys Lys
2210 2215 2220

Lys Leu Ser Pro Pro Ser Ile Ser Ile Asp Pro Pro Glu Ser Gln Gly
2225 2230 2235 2240

Ser Arg Pro Pro Cys Ser Pro Gly Val Cys Leu Arg Arg Arg Ala Pro
2245 2250 2255

Ala Ser Asp Ser Lys Asp Pro Ser Val Ser Ser Pro Leu Asp Ser Thr
2260 2265 2270

Ala Ala Ser Pro Ser Pro Lys Lys Asp Thr Leu Ser Leu Ser Gly Leu
2275 2280 2285

Ser Ser Asp Pro Thr Asp Met Asp Pro
2290 2295

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2304 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Leu Pro His Arg Val Pro Arg Cys Val Arg Thr Pro Pro Leu Arg
1 5 10 15

Gly Ser Ala Arg Pro Ser Ser Asp Pro Pro Gly Pro Arg Leu Ala Arg
20 25 30

Gly Trp Thr Arg Arg Arg Met Glu Arg Ala Pro Arg Ser Arg Asp Ser

35 40 45

Pro Val Ala Ser Arg Ser Ser Thr Thr Cys Pro Gly Pro Gly Ala Ala
50 55 60

Gly Ala Gly Ser Thr Glu Lys Asp Pro Gly Ser Ala Asp Ser Glu Ala
65 70 75 80

Glu Gly Leu Pro Tyr Pro Ala Leu Ala Pro Val Val Phe Phe Tyr Leu
85 90 95

Ser Gln Asp Ser Arg Pro Arg Ser Trp Cys Leu Arg Thr Val Cys Asn
100 105 110

Pro Trp Phe Glu Arg Val Ser Met Leu Val Ile Leu Leu Asn Cys Val
115 120 125

Thr Leu Gly Met Phe Arg Pro Cys Glu Asp Ile Ala Cys Asp Ser Gln
130 135 140

Arg Cys Arg Ile Leu Gln Ala Phe Asp Asp Phe Ile Phe Ala Phe Phe
145 150 155 160

Ala Val Glu Met Val Val Lys Met Val Ala Leu Gly Ile Phe Gly Lys
165 170 175

Lys Cys Tyr Leu Gly Asp Thr Trp Asn Arg Leu Asp Phe Phe Ile Val
180 185 190

Ile Ala Gly Met Leu Glu Tyr Ser Leu Asp Leu Gln Asn Val Ser Phe
195 200 205

Ser Ala Val Arg Thr Val Arg Val Leu Arg Pro Leu Arg Ala Ile Asn
210 215 220

Arg Val Pro Ser Met Arg Ile Leu Val Thr Leu Leu Leu Asp Thr Leu
225 230 235 240

```
Pro Met Leu Gly Asn Val Leu Leu Leu Cys Phe Phe Val Phe Phe Ile
                245                 250                 255

Phe Gly Ile Val Gly Val Gln Leu Trp Ala Gly Leu Leu Arg Asn Arg
            260                 265                 270

Cys Phe Leu Pro Glu Asn Phe Ser Leu Pro Leu Ser Val Asp Leu Glu
        275                 280                 285

Pro Tyr Tyr Gln Thr Glu Asn Glu Asp Glu Ser Pro Phe Ile Cys Ser
    290                 295                 300

Gln Pro Arg Glu Asn Gly Met Arg Ser Cys Arg Ser Val Pro Thr Leu
305                 310                 315                 320

Arg Gly Glu Gly Gly Gly Gly Pro Pro Cys Ser Leu Asp Tyr Glu Thr
                325                 330                 335

Tyr Asn Ser Ser Ser Asn Thr Thr Cys Val Asn Trp Asn Gln Tyr Tyr
            340                 345                 350

Thr Asn Cys Ser Ala Gly Glu His Asn Pro Phe Lys Gly Ala Ile Asn
        355                 360                 365

Phe Asp Asn Ile Gly Tyr Ala Trp Ile Ala Ile Phe Gln Val Ile Thr
    370                 375                 380

Leu Glu Gly Trp Val Asp Ile Met Tyr Phe Val Met Asp Ala His Ser
385                 390                 395                 400

Phe Tyr Asn Phe Ile Tyr Phe Ile Leu Leu Ile Ile Val Gly Ser Phe
                405                 410                 415

Phe Met Ile Asn Leu Cys Leu Val Val Ile Ala Thr Gln Phe Ser Glu
            420                 425                 430
```

Thr Lys Gln Arg Glu Ser Gln Leu Met Arg Glu Gln Arg Val Arg Phe
435 440 445

Leu Ser Asn Ala Ser Thr Leu Ala Ser Phe Ser Glu Pro Gly Ser Cys
450 455 460

Tyr Glu Glu Leu Leu Lys Tyr Leu Val Tyr Ile Leu Arg Lys Ala Ala
465 470 475 480

Arg Arg Leu Ala Gln Val Ser Arg Ala Ile Gly Val Arg Ala Gly Leu
485 490 495

Leu Ser Ser Pro Val Ala Arg Ser Gly Gln Glu Pro Gln Pro Ser Gly
500 505 510

Ser Cys Thr Arg Ser His Arg Arg Leu Ser Val His His Leu Val His
515 520 525

His His His His His His His His Tyr His Leu Gly Asn Gly Thr Leu
530 535 540

Arg Val Pro Arg Ala Ser Pro Glu Ile Gln Asp Arg Asp Ala Asn Gly
545 550 555 560

Ser Arg Arg Leu Met Leu Pro Pro Pro Ser Thr Pro Thr Pro Ser Gly
565 570 575

Gly Pro Pro Arg Gly Ala Glu Ser Val His Ser Phe Tyr His Ala Asp
580 585 590

Cys His Leu Glu Pro Val Arg Cys Gln Ala Pro Pro Pro Arg Cys Pro
595 600 605

Ser Glu Ala Ser Gly Arg Thr Val Gly Ser Gly Lys Val Tyr Pro Thr
610 615 620

Val His Thr Ser Pro Pro Pro Glu Ile Leu Lys Asp Lys Ala Leu Val

625 630 635 640

Glu Val Ala Pro Ser Pro Gly Pro Pro Thr Leu Thr Ser Phe Asn Ile
645 650 655

Pro Pro Gly Pro Phe Ser Ser Met His Lys Leu Leu Glu Thr Gln Ser
660 665 670

Thr Gly Ala Cys His Ser Ser Cys Lys Ile Ser Ser Pro Cys Ser Lys
675 680 685

Ala Asp Ser Gly Ala Cys Gly Pro Asp Ser Cys Pro Tyr Cys Ala Arg
690 695 700

Thr Gly Ala Gly Glu Pro Glu Ser Ala Asp His Val Met Pro Asp Ser
705 710 715 720

Asp Ser Glu Ala Val Tyr Glu Phe Thr Gln Asp Ala Gln His Ser Asp
725 730 735

Leu Arg Asp Pro His Ser Arg Arg Arg Gln Arg Ser Leu Gly Pro Asp
740 745 750

Ala Glu Pro Ser Ser Val Leu Ala Phe Trp Arg Leu Ile Cys Asp Thr
755 760 765

Phe Arg Lys Ile Val Asp Ser Lys Tyr Phe Gly Arg Gly Ile Met Ile
770 775 780

Ala Ile Leu Val Asn Thr Leu Ser Met Gly Ile Glu Tyr His Glu Gln
785 790 795 800

Pro Glu Glu Leu Thr Asn Ala Leu Glu Ile Ser Asn Ile Val Phe Thr
805 810 815

Ser Leu Phe Ala Leu Glu Met Leu Leu Lys Leu Leu Val Tyr Gly Pro
820 825 830

Phe Gly Tyr Ile Lys Asn Pro Tyr Asn Ile Phe Asp Gly Val Ile Val
835 840 845

Val Ile Ser Val Trp Glu Ile Val Gly Gln Gln Gly Gly Gly Leu Ser
850 855 860

Val Leu Arg Thr Phe Arg Leu Met Arg Val Leu Lys Leu Val Arg Phe
865 870 875 880

Leu Pro Ala Leu Gln Arg Gln Leu Val Val Leu Met Lys Thr Met Asp
885 890 895

Asn Val Ala Thr Phe Cys Met Leu Leu Met Leu Phe Ile Phe Ile Phe
900 905 910

Ser Ile Leu Gly Met His Leu Phe Gly Cys Lys Phe Ala Ser Glu Arg
915 920 925

Asp Gly Asp Thr Leu Pro Asp Arg Lys Asn Phe Asp Ser Leu Leu Trp
930 935 940

Ala Ile Val Thr Val Phe Gln Ile Leu Thr Gln Glu Asp Trp Asn Lys
945 950 955 960

Val Leu Tyr Asn Gly Met Ala Ser Thr Ser Ser Trp Ala Ala Leu Tyr
965 970 975

Phe Ile Ala Leu Met Thr Phe Gly Asn Tyr Val Leu Phe Asn Leu Leu
980 985 990

Val Ala Ile Leu Val Glu Gly Phe Gln Ala Glu Gly Asp Ala Thr Lys
995 1000 1005

Ser Glu Ser Glu Pro Asp Phe Phe Ser Pro Ser Val Asp Gly Asp Gly
1010 1015 1020

Asp Arg Lys Lys Arg Leu Ala Leu Val Ala Leu Gly Glu His Ala Glu
1025 1030 1035 1040

Leu Arg Lys Ser Leu Leu Pro Pro Leu Ile Ile His Thr Ala Ala Thr
1045 1050 1055

Pro Met Ser His Pro Lys Ser Ser Ser Thr Gly Val Gly Glu Ala Leu
1060 1065 1070

Gly Ser Gly Ser Arg Arg Thr Ser Ser Ser Gly Ser Ala Glu Pro Gly
1075 1080 1085

Ala Ala His His Glu Met Lys Cys Pro Pro Ser Ala Arg Ser Ser Pro
1090 1095 1100

His Ser Pro Trp Ser Ala Ala Ser Ser Trp Thr Ser Arg Arg Ser Ser
1105 1110 1115 1120

Arg Asn Ser Leu Gly Arg Ala Pro Ser Leu Lys Arg Arg Ser Pro Ser
1125 1130 1135

Gly Glu Arg Arg Ser Leu Leu Ser Gly Glu Gly Gln Glu Ser Gln Asp
1140 1145 1150

Glu Glu Glu Ser Ser Glu Glu Asp Arg Ala Ser Pro Ala Gly Ser Asp
1155 1160 1165

His Arg His Arg Gly Ser Leu Glu Arg Glu Ala Lys Ser Ser Phe Asp
1170 1175 1180

Leu Pro Asp Thr Leu Gln Val Pro Gly Leu His Arg Thr Ala Ser Gly
1185 1190 1195 1200

Arg Ser Ser Ala Ser Glu His Gln Asp Cys Asn Gly Lys Ser Ala Ser
1205 1210 1215

Gly Arg Leu Ala Arg Thr Leu Arg Thr Asp Asp Pro Gln Leu Asp Gly

1220 1225 1230

Asp Asp Asp Asn Asp Glu Gly Asn Leu Ser Lys Gly Glu Arg Ile Gln
1235 1240 1245

Ala Trp Val Arg Ser Arg Leu Pro Ala Cys Cys Arg Glu Arg Asp Ser
1250 1255 1260

Trp Ser Ala Tyr Ile Phe Pro Pro Gln Ser Arg Phe Arg Leu Leu Cys
1265 1270 1275 1280

His Arg Ile Ile Thr His Lys Met Phe Asp His Val Val Leu Val Ile
1285 1290 1295

Ile Phe Leu Asn Cys Ile Thr Ile Ala Met Glu Arg Pro Lys Ile Asp
1300 1305 1310

Pro His Ser Ala Glu Arg Ile Phe Leu Thr Leu Ser Asn Tyr Ile Phe
1315 1320 1325

Thr Ala Val Phe Leu Ala Glu Met Thr Val Lys Val Val Ala Leu Gly
1330 1335 1340

Trp Cys Phe Gly Glu Gln Ala Tyr Leu Arg Ser Ser Trp Asn Val Leu
1345 1350 1355 1360

Asp Gly Leu Leu Val Leu Ile Ser Val Ile Asp Ile Leu Val Ser Met
1365 1370 1375

Val Ser Asp Ser Gly Thr Lys Ile Leu Gly Met Leu Arg Val Leu Arg
1380 1385 1390

Leu Leu Arg Thr Leu Arg Pro Leu Arg Val Ile Ser Arg Ala Gln Gly
1395 1400 1405

Leu Lys Leu Val Val Glu Thr Leu Met Ser Ser Leu Lys Pro Ile Gly
1410 1415 1420

Asn Ile Val Val Ile Cys Cys Ala Phe Phe Ile Ile Phe Gly Ile Leu
1425 1430 1435 1440

Gly Val Gln Leu Phe Lys Gly Lys Phe Phe Val Cys Gln Gly Glu Asp
1445 1450 1455

Thr Arg Asn Ile Thr Asn Lys Ser Asp Cys Ala Glu Ala Ser Tyr Arg
1460 1465 1470

Trp Val Arg His Lys Tyr Asn Phe Asp Asn Leu Gly Gln Ala Leu Met
1475 1480 1485

Ser Leu Phe Val Leu Ala Ser Lys Asp Gly Trp Val Asp Ile Met Tyr
1490 1495 1500

Asp Gly Leu Asp Ala Val Gly Val Asp Gln Gln Pro Ile Met Asn His
1505 1510 1515 1520

Asn Pro Trp Met Leu Leu Tyr Phe Ile Ser Phe Leu Leu Ile Val Ala
1525 1530 1535

Phe Phe Val Leu Asn Met Phe Val Gly Val Val Val Glu Asn Phe His
1540 1545 1550

Lys Cys Arg Gln His Gln Glu Glu Glu Glu Ala Arg Arg Arg Glu Glu
1555 1560 1565

Lys Arg Leu Arg Arg Leu Glu Lys Lys Arg Arg Ser Lys Glu Lys Gln
1570 1575 1580

Met Ala Asp Leu Met Leu Asp Asp Val Ile Ala Ser Gly Ser Ser Ala
1585 1590 1595 1600

Ser Ala Ala Ser Glu Ala Gln Cys Lys Pro Tyr Tyr Ser Asp Tyr Ser
1605 1610 1615

Arg Phe Arg Leu Leu Val His His Leu Cys Thr Ser His Tyr Leu Asp
1620 1625 1630

Leu Phe Ile Thr Gly Val Ile Gly Leu Asn Val Val Thr Met Ala Met
1635 1640 1645

Glu His Tyr Gln Gln Pro Gln Ile Leu Asp Glu Ala Leu Lys Ile Cys
1650 1655 1660

Asn Tyr Ile Phe Thr Val Ile Phe Val Phe Glu Ser Val Phe Lys Leu
1665 1670 1675 1680

Val Ala Phe Ala Phe Arg Arg Phe Phe Gln Asp Arg Trp Asn Gln Leu
1685 1690 1695

Asp Leu Ala Ile Val Leu Leu Ser Ile Met Gly Ile Thr Leu Glu Glu
1700 1705 1710

Ile Glu Val Asn Leu Ser Leu Pro Ile Asn Pro Thr Ile Ile Arg Ile
1715 1720 1725

Met Arg Val Leu Arg Ile Ala Arg Val Leu Lys Leu Leu Lys Met Ala
1730 1735 1740

Val Gly Met Arg Ala Leu Leu His Thr Val Met Gln Ala Leu Pro Gln
1745 1750 1755 1760

Val Gly Asn Leu Gly Leu Leu Phe Met Leu Leu Phe Phe Ile Phe Ala
1765 1770 1775

Ala Leu Gly Val Glu Leu Phe Gly Asp Leu Glu Cys Asp Glu Thr His
1780 1785 1790

Pro Cys Glu Gly Leu Gly Arg His Ala Thr Phe Arg Asn Phe Gly Met
1795 1800 1805

Ala Phe Leu Thr Leu Phe Arg Val Ser Thr Gly Asp Asn Trp Asn Gly

1810 1815 1820

Ile Met Lys Asp Pro Ser Arg Asp Cys Asp Gln Glu Ser Thr Cys Tyr
1825 1830 1835 1840

Asn Thr Val Ile Ser Pro Ile Tyr Phe Val Ser Phe Val Leu Thr Ala
1845 1850 1855

Gln Phe Val Leu Val Asn Val Val Ile Ala Val Leu Met Lys His Leu
1860 1865 1870

Glu Glu Ser Asn Lys Glu Ala Lys Glu Glu Ala Glu Leu Glu Ala Glu
1875 1880 1885

Leu Glu Leu Glu Met Lys Thr Leu Ser Pro Gln Pro His Ser Pro Leu
1890 1895 1900

Gly Ser Pro Phe Leu Trp Pro Gly Val Glu Gly Val Asn Ser Thr Asp
1905 1910 1915 1920

Ser Pro Lys Pro Gly Ala Pro His Thr Thr Ala His Ile Gly Ala Ala
1925 1930 1935

Ser Gly Phe Ser Leu Glu His Pro Thr Met Val Pro His Pro Glu Glu
1940 1945 1950

Val Pro Val Pro Leu Gly Pro Asp Leu Leu Thr Val Arg Lys Ser Gly
1955 1960 1965

Val Ser Arg Thr His Ser Leu Pro Asn Asp Ser Tyr Met Cys Arg Asn
1970 1975 1980

Gly Ser Thr Ala Glu Arg Ser Leu Gly His Arg Gly Trp Gly Leu Pro
1985 1990 1995 2000

Lys Ala Gln Ser Gly Ser Ile Leu Ser Val His Ser Gln Pro Ala Asp
2005 2010 2015

Thr Ser Cys Ile Leu Gln Leu Pro Lys Asp Val His Tyr Leu Leu Gln
2020 2025 2030

Pro His Gly Ala Pro Thr Trp Gly Ala Ile Pro Lys Leu Pro Pro Pro
2035 2040 2045

Gly Arg Ser Pro Leu Ala Gln Arg Pro Leu Arg Arg Gln Ala Ala Ile
2050 2055 2060

Arg Thr Asp Ser Leu Asp Val Gln Gly Leu Gly Ser Arg Glu Asp Leu
2065 2070 2075 2080

Leu Ser Glu Val Ser Gly Pro Ser Cys Pro Leu Thr Arg Ser Ser Ser
2085 2090 2095

Phe Trp Gly Gly Ser Ser Ile Gln Val Gln Gln Arg Ser Gly Ile Gln
2100 2105 2110

Ser Lys Val Ser Lys His Ile Arg Leu Pro Ala Pro Cys Pro Gly Leu
2115 2120 2125

Glu Pro Ser Trp Ala Lys Asp Pro Pro Glu Thr Arg Ser Ser Leu Glu
2130 2135 2140

Leu Asp Thr Glu Leu Ser Trp Ile Ser Gly Asp Leu Leu Pro Ser Ser
2145 2150 2155 2160

Gln Glu Glu Pro Leu Phe Pro Arg Asp Leu Lys Lys Cys Tyr Ser Val
2165 2170 2175

Glu Thr Gln Ser Cys Arg Arg Arg Pro Gly Phe Trp Leu Asp Glu Gln
2180 2185 2190

Arg Arg His Ser Ile Ala Val Ser Cys Leu Asp Ser Gly Ser Gln Pro
2195 2200 2205

Arg Leu Cys Pro Ser Pro Ser Ser Leu Gly Gly Gln Pro Leu Gly Gly
2210 2215 2220

Pro Gly Ser Arg Pro Lys Lys Lys Leu Ser Pro Pro Ser Ile Ser Ile
2225 2230 2235 2240

Asp Pro Pro Glu Ser Gln Gly Ser Arg Pro Pro Cys Ser Pro Gly Val
2245 2250 2255

Cys Leu Arg Arg Arg Ala Pro Ala Ser Asp Ser Lys Asp Pro Ser Val
2260 2265 2270

Ser Ser Pro Leu Asp Ser Thr Ala Ala Ser Pro Ser Pro Lys Lys Asp
2275 2280 2285

Thr Leu Ser Leu Ser Gly Leu Ser Ser Asp Pro Thr Asp Met Asp Pro
2290 2295 2300

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Arg Ile Met Arg Val Leu Arg Ile Ala Arg Val Leu Lys Leu Leu Lys
1 5 10 15

Met Ala Val Gly Met Arg Ala
20

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Arg Leu Phe Arg Val Met Arg Leu Ile Lys Leu Leu Ser Arg Ala Glu
1 5 10 15

Gly Val

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Arg	Leu	Phe	Arg	Val	Met	Arg	Leu	Val	Lys	Leu	Leu	Ser	Arg	Gly	Glu
1				5					10						15

Gly Ile

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Arg	Leu	Phe	Arg	Ala	Ala	Arg	Leu	Ile	Lys	Leu	Leu	Arg	Gln	Gly	Tyr
1				5					10						15

Thr Ile

What is claimed is:

1. A isolated or substantially purified nucleic acid encoding a protein comprising at least one domain of a T-type calcium channel α subunit.

2. The nucleic acid of claim 1, which encodes an entire T-type calcium channel α subunit.

3. The nucleic acid of claim 2, wherein said protein comprises SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6 or a derivative of any of said sequences.

4. The nucleic acid of claim 1, wherein said protein comprises SEQ ID NO:7.

5. The nucleic acid of claim 2, wherein said protein gates from about -45 mV to about -30 mV in 2 mM Ba^{2+}.

6. The nucleic acid of claim 2, wherein said protein exhibits a tail current of from about 2 ms to about 10 ms following repolarization to a membrane potential from about –80 mV to about –60 mV in a solution with a barium concentration of from about 10 mM to about 40 mM.

7. The nucleic acid of claim 2, wherein said protein exhibits a single channel conductance of from 7 pS to about 10 pS in a solution with a barium ion concentration of about 100 mM.

8. A isolated or substantially purified nucleic acid hybridizing to SEQ ID NO:2 or SEQ ID NO:4 under high stringency.

9. A isolated or substantially purified DNA hybridizing to the nucleic acid of claim 8.

10. The DNA of claim 9 comprising a sequence encoding a T-type calcium channel.

11. A vector comprising the nucleic acid of claim 1.

12. A cell into which the vector of claim 11 has been introduced.

13. The cell of claim 12, wherein said nucleic acid is expressed to produce a protein.

14. A method of identifying a drug which affects T-type calcium channels, said method comprising expressing a T-type calcium channel in a cell, exposing said cell to a putative drug, and measuring the calcium flux through the membrane of said cell in response to a change in membrane potential.

15. The method of claim 14, wherein said calcium flux is assayed by using a calcium-sensitive labile dye within said cell.

16. The method of claim 14, wherein said calcium flux is assayed by measuring the electrophysiological properties of said cell.

17. The method of claim 14, wherein said calcium channel comprises SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6, or a derivative of any of said sequences.

18. The method of claim 14, wherein said calcium channel comprises SEQ ID NO:7.

19. An isolated or substantially purified antibody molecule recognizing an epitope on a T-type calcium channel protein.

ABSTRACT

The present invention provides an isolated or substnatially purified nucleic acid encoding a protein comprising at least one domain of a T-type calcium channel α subunit and cells expressing such nucleic acids. The present invention also provides isolated or substantially purified T-type calcium channels and an isolated or substantially purified antibody molecule recognizing an epitope on a T-type calcium channel protein. Additionally, the present invention provides an assay for identifying potential drugs affecting T-type calcium channels by exposing cells expressing a nucleic acid encoding a T-type calcium channel to a putative drug and then measuring the calcium flux in response to a change in membrane potential.

```
          ATGCTCCCCCACCGGGTCCCCCGTTGCGTGAGGACACCTCCTCTGAGGGGCTCCGCTCGCCCCTCT
  1         M  L  P  H  R  V  P  R  C  V  R  T  P  P  L  R  G  S  A  R  P  S

 66 TCGGACCCCCCGGGGCCCCGGCTGGCCAGAGGATGGACGAGGAGGAGGATGGAGCGGGCGCCGAGGAGTCGGGAC
 23  S  D  P  P  G  P  R  L  A  R  G  W  T  R  R  R  M  E  R  A  P  R  S  R  D

141 AGCCCCGTAGCTTCACGCAGCTCAACGACCTGTCCGGGGCCGGGGGCGGCAGGGGCCGGGTCGACGGAAAAGGAC
 48  S  P  V  A  S  R  S  S  T  T  C  P  G  P  G  A  A  G  A  G  S  T  E  K  D

216 CCGGGCAGCGCGGACTCCGAGGCGGAGGGGCTGCCGTACCCGGCGCTAGCCCCGGTGGTTTTCTTCTACTTGAGC
 73  P  G  S  A  D  S  E  A  E  G  L  P  Y  P  A  L  A  P  V  V  F  F  Y  L  S

                                                    ********************************
291 CAGGACAGCCGCCCGCGGAGCTGGTGTCTCCGCACGGTCTGTAACCCGTGGTTCGAGCGAGTCAGTATGCTGGTC
 98  Q  D  S  R  P  R  S  W  C  L  R  T  V  C  N  P  W  F  E  R  V  S  M  L  V

    ******IS1*****************************
366 ATTCTTCTCAACTGTGTGACTCTGGGTATGTTCAGGCCGTGTGAGGACATTGCCTGTGACTCCCAGCGCTGCCGG
123  I  L  L  N  C  V  T  L  G  M  F  R  P  C  E  D  I  A  C  D  S  Q  R  C  R

          ****************************************IS2**************************
441 ATCCTGCAGGCCTTCGATGACTTCATCTTTGCCTTCTTTGCTGTGGAAATGGTGGTGAAGATGGTGGCCTTGGGC
148  I  L  Q  A  F  D  D  F  I  F  A  F  F  A  V  E  M  V  V  K  M  V  A  L  G

    ***                                 ***********************IS3*****************
516 ATCTTTGGGAAGAAATGTTACCTGGGAGACACTTGGAACCGGCTTGACTTTTTCATTGTCATTGCAGGGATGCTG
173  I  F  G  K  K  C  Y  L  G  D  T  W  N  R  L  D  F  F  I  V  I  A  G  M  L

    ***************                     **********************IS4*****************
591 GAGTATTCGCTGGACCTGCAGAACGTCAGCTTCTCCGCAGTCAGGACAGTCCGTGTGCTGCGACCGCTCAGGGCC
198  E  Y  S  L  D  L  Q  N  V  S  F  S  A  V  R  T  V  R  V  L  R  P  L  R  A

    ****************************                                  ******************
666 ATTAACCGGGTGCCCAGCATGCGCATTCTCGTCACATTACTGCTGGACACCTTGCCTATGCTGGGCAACGTCCTG
223  I  N  R  V  P  S  M  R  I  L  V  T  L  L  L  D  T  L  P  M  L  G  N  V  L

    ********************************IS5**********************************
741 CTGCTCTGTTTCTTCGTCTTTTTCATCTTTGGCATCGTGGGCGTCCAGCTGTGGGCAGGACTGCTTCGCAACCGG
248  L  L  C  F  F  V  F  F  I  F  G  I  V  G  V  Q  L  W  A  G  L  L  R  N  R

816 TGCTTCCTCCCCGAGAACTTCAGCCTCCCCCTGAGCGTGGACCTGGAGCCTTATTACCAGACAGAGAATGAGGAC
273  C  F  L  P  E  N  F  S  L  P  L  S  V  D  L  E  P  Y  Y  Q  T  E  N  E  D

891 GAGAGCCCCTTCATCTGCTCTCAGCCTCGGGAGAATGGCATGAGATCCTGCAGGAGTGTGCCCACACTGCGTGGG
298  E  S  P  F  I  C  S  Q  P  R  E  N  G  M  R  S  C  R  S  V  P  T  L  R  G
```

Figure 1A

```
966 GAAGGCGGTGGTGGCCCACCCTGCAGTCTGGACTATGAGACCTATAACAGTTCCAGCAACACCACCTGTGTCAAC
323  E  G  G  G  G  P  P  C  S  L  D  Y  E  T  Y  N  S  S  S  N  T  T  C  V  N

                                                          ************
1041 TGGAACCAGTACTATACCAACTGCTCTGCGGGCGAGCACAACCCCTTCAAAGGCGCCATCAACTTTGACAACATT
348  W  N  Q  Y  Y  T  N  C  S  A  G  E  H  N  P  F  K  G  A  I  N  F  D  N  I

     **************I P Loop****************************************
1116 GGCTATGCCTGGATCGCCATCTTCCAGGTCATCACACTGGAGGGCTGGGTCGACATCATGTACTTCGTAATGGAC
373  G  Y  A  W  I  A  I  F  Q  V  I  T  L  E  G  W  V  D  I  M  Y  F  V  M  D

                  ********************************IS6*********************************
1191 GCTCACTCCTTCTACAACTTCATCTACTTCATTCTTCTCATCATCGTGGGCTCCTTCTTCATGATCAACCTGTGC
398  A  H  S  F  Y  N  F  I  Y  F  I  L  L  I  I  V  G  S  F  F  M  I  N  L  C

1266 CTGGTGGTGATTGCCACGCAGTTCTCCGAGACCAAACAGCGGGAGAGTCAGCTGATGCGGGAGCAGCGTGTACGA
423  L  V  V  I  A  T  Q  F  S  E  T  K  Q  R  E  S  Q  L  M  R  E  Q  R  V  R

1341 TTCCTGTCCAATGCTAGCACCCTGGCAAGCTTCTCTGAGCCAGGCAGCTGCTATGAGGAGCTACTCAAGTACCTG
448  F  L  S  N  A  S  T  L  A  S  F  S  E  P  G  S  C  Y  E  E  L  L  K  Y  L

1416 GTGTACATCCTCCGAAAAGCAGCCCGAAGGCTGGCCCAGGTCTCTAGGGCTATAGGCGTGCGGGCTGGGCTGCTC
473  V  Y  I  L  R  K  A  A  R  R  L  A  Q  V  S  R  A  I  G  V  R  A  G  L  L

1491 AGCAGCCCAGTGGCCCGTAGTGGGCAGGAGCCCCAGCCCAGTGGCAGCTGCACTCGCTCACACCGTCGTCTGTCT
498  S  S  P  V  A  R  S  G  Q  E  P  Q  P  S  G  S  C  T  R  S  H  R  R  L  S

1566 GTCCACCACCTGGTCCACCACCATCACCACCACCATCACCACTACCACCTGGGTAATGGGACGCTCAGAGTTCCC
523  V  H  H  L  V  H  H  H  H  H  H  H  H  H  Y  H  L  G  N  G  T  L  R  V  P

1641 CGGGCCAGCCCAGAGATCCAGGACAGGGATGCCAATGGGTCTCGCCGGCTCATGCTACCACCACCCTCTACACCC
548  R  A  S  P  E  I  Q  D  R  D  A  N  G  S  R  R  L  M  L  P  P  P  S  T  P

1716 ACTCCCTCTGGGGGCCCTCCGAGGGGTGCGGAGTCTGTACACAGCTTCTACCATGCTGACTGCCACTTGGAGCCA
573  T  P  S  G  G  P  P  R  G  A  E  S  V  H  S  F  Y  H  A  D  C  H  L  E  P

1791 GTCCGTTGCCAGGCACCCCCTCCCAGATGCCCATCGGAGGCATCTGGTAGGACTGTGGGTAGTGGGAAGGTGTAC
598  V  R  C  Q  A  P  P  P  R  C  P  S  E  A  S  G  R  T  V  G  S  G  K  V  Y

1866 CCCACTGTGCATACCAGCCCTCCACCAGAGATACTGAAGGATAAAGCACTAGTGGAGGTGGCCCCCAGCCCTGGG
623  P  T  V  H  T  S  P  P  P  E  I  L  K  D  K  A  L  V  E  V  A  P  S  P  G

1941 CCCCCCACCCTCACCAGCTTCAACATCCCACCTGGGCCCTTCAGCTCCATGCACAAGCTCCTGGAGACACAGAGT
648  P  P  T  L  T  S  F  N  I  P  P  G  P  F  S  S  M  H  K  L  L  E  T  Q  S

2016 ACGGGAGCCTGCCATAGCTCCTGCAAAATCTCCAGCCCTTGCTCCAAGGCAGACAGTGGAGCCTGCGGGCCGGAC
673  T  G  A  C  H  S  S  C  K  I  S  S  P  C  S  K  A  D  S  G  A  C  G  P  D

2091 AGTTGTCCCTACTGTGCCCGGACAGGAGCAGGAGAGCCAGAGTCCGCTGACCATGTCATGCCTGACTCAGACAGC
698  S  C  P  Y  C  A  R  T  G  A  G  E  P  E  S  A  D  H  V  M  P  D  S  D  S

2166 GAGGCTGTGTATGAGTTCACACAGGACGCTCAGCACAGTGACCTCCGGGATCCCCACAGCCGGCGGCGACAGCGG
723  E  A  V  Y  E  F  T  Q  D  A  Q  H  S  D  L  R  D  P  H  S  R  R  R  Q  R

2241 AGCCTGGGCCCAGATGCAGAGCCTAGTTCTGTGCTGGCTTTCTGGAGGCTGATCTGTGACACATTCCGGAAGATC
748  S  L  G  P  D  A  E  P  S  S  V  L  A  F  W  R  L  I  C  D  T  F  R  K  I
```

Figure 1B

```
                    ************************************IIS1**************************
2316 GTAGATAGCAAATACTTTGGCCGGGGAATCATGATCGCCATCCTGGTCAATACACTCAGCATGGGCATCGAGTAC
 773  V  D  S  K  Y  F  G  R  G  I  M  I  A  I  L  V  N  T  L  S  M  G  I  E  Y

     ***                                          ************************************
2391 CACGAGCAGCCCGAGGAGCTCACCAACGCCCTGGAAATCAGCAACATCGTCTTCACCAGCCTCTTCGCCTTGGAG
 798  H  E  Q  P  E  E  L  T  N  A  L  E  I  S  N  I  V  F  T  S  L  F  A  L  E

     ****IIS2**********************                  ******************IIS3**
2466 ATGCTGCTGAAACTGCTTGTCTACGGTCCCTTTGGCTACATTAAGAATCCCTACAACATCTTTGATGGTGTCATT
 823  M  L  L  K  L  L  V  Y  G  P  F  G  Y  I  K  N  P  Y  N  I  F  D  G  V  I

     **********************************           **********************
2541 GTGGTCATCAGTGTGTGGGAGATTGTGGGCCAGCAGGGAGGTGGCCTGTCGGTGCTGCGGACCTTCCGCCTGATG
 848  V  V  I  S  V  W  E  I  V  G  Q  Q  G  G  G  L  S  V  L  R  T  F  R  L  M

     ******************IIS4*****************************                  ***
2616 CGGGTGCTGAAGCTGGTGCGCTTCCTGCCGGCCCTGCAGCGCCAGCTCGTGGTGCTCATGAAGACCATGGACAAC
 873  R  V  L  K  L  V  R  F  L  P  A  L  Q  R  Q  L  V  V  L  M  K  T  M  D  N

     **********************************IIS5************************************
2691 GTGGCCACCTTCTGCATGCTCCTCATGCTGTTCATCTTCATCTTCAGCATCCTGGGCATGCATCTCTTTGGTTGC
 898  V  A  T  F  C  M  L  L  M  L  F  I  F  I  F  S  I  L  G  M  H  L  F  G  C

     ***                                          ****************************
2766 AAGTTCGCATCTGAACGGGATGGGGACACGTTGCCAGACCGGAAGAATTTCGACTCCCTGCTCTGGGCCATCGTC
 923  K  F  A  S  E  R  D  G  D  T  L  P  D  R  K  N  F  D  S  L  L  W  A  I  V

     **********II Pore Loop**********************                  ******
2841 ACTGTCTTTCAGATTCTGACTCAGGAAGACTGGAATAAAGTCCTCTACAACGGCATGGCCTCCACATCGTCTTGG
 948  T  V  F  Q  I  L  T  Q  E  D  W  N  K  V  L  Y  N  G  M  A  S  T  S  S  W

     **********************************IIS6************************************
2916 GCTGCTCTTTACTTCATCGCCCTCATGACTTTTGGCAACTATGTGCTCTTTAACCTGCTGGTGGCCATTCTTGTG
 973  A  A  L  Y  F  I  A  L  M  T  F  G  N  Y  V  L  F  N  L  L  V  A  I  L  V

     *********
2991 GAAGGATTCCAGGCAGAGGGAGATGCCACCAAGTCTGAGTCAGAGCCTGATTTCTTTTCGCCCAGTGTGGATGGT
 998  E  G  F  Q  A  E  G  D  A  T  K  S  E  S  E  P  D  F  F  S  P  S  V  D  G

3066 GATGGGGACAGAAAGAAGCGCTTGGCCCTGGTGGCTTTGGGAGAACACGCGGAACTACGAAAGAGCCTTTTGCCA
1023  D  G  D  R  K  K  R  L  A  L  V  A  L  G  E  H  A  E  L  R  K  S  L  L  P

3141 CCCCTCATCATCCATACGGCTGCGACACCAATGTCACACCCCAAGAGCTCCAGCACAGGTGTGGGGGAAGCACTG
1048  P  L  I  I  H  T  A  A  T  P  M  S  H  P  K  S  S  S  T  G  V  G  E  A  L

3216 GGCTCTGGCTCTCGACGTACCAGTAGCAGTGGGTCCGCTGAGCCTGGAGCTGCCCACCATGAGATGAAATGTCCG
1073  G  S  G  S  R  R  T  S  S  S  G  S  A  E  P  G  A  A  H  H  E  M  K  C  P

3291 CCAAGTGCCCGCAGCTCCCCGCACAGTCCCTGGAGTGCGGCAAGCAGCTGGACCAGCAGGCGCTCCAGCAGGAAC
1098  P  S  A  R  S  S  P  H  S  P  W  S  A  A  S  S  W  T  S  R  R  S  S  R  N

3366 AGCCTGGGCCGGGCCCCCAGCCTAAAGCGGAGGAGCCCGAGCGGGGAGCGGAGGTCCCTGCTGTCTGGAGAGGGC
1123  S  L  G  R  A  P  S  L  K  R  R  S  P  S  G  E  R  R  S  L  L  S  G  E  G
```

Figure 1C

```
3441 CAGGAGAGTCAGGATGAGGAGGAAAGTTCAGAAGAGGACCGGGCCAGCCCAGCAGGCAGTGACCATCGCCACAGG
1148  Q  E  S  Q  D  E  E  E  S  S  E  E  D  R  A  S  P  A  G  S  D  H  R  H  R

3516 GGTTCCTTGGAACGTGAGGCCAAGAGTTCCTTTGACCTGCCTGACACTCTGCAGGTGCCGGGGCTGCACCGCACA
1173  G  S  L  E  R  E  A  K  S  S  F  D  L  P  D  T  L  Q  V  P  G  L  H  R  T

3591 GCCAGCGGCCGGAGCTCTGCCTCTGAGCACCAAGACTGTAATGGCAAGTCGGCTTCAGGGCGTTTGGCCCGCACC
1198  A  S  G  R  S  S  A  S  E  H  Q  D  C  N  G  K  S  A  S  G  R  L  A  R  T

3666 CTGAGGACTGATGACCCCCAACTGGATGGGGATGATGACAATGATGAGGGAAATCTGAGCAAAGGGGAACGCATA
1223  L  R  T  D  D  P  Q  L  D  G  D  D  D  N  D  E  G  N  L  S  K  G  E  R  I

3741 CAAGCCTGGGTCAGATCCCGGCTTCCTGCCTGTTGCCGAGAGCGAGATTCCTGGTCGGCCTATATCTTTCCTCCT
1248  Q  A  W  V  R  S  R  L  P  A  C  C  R  E  R  D  S  W  S  A  Y  I  F  P  P

                                                  **********IIIS1*****
3816 CAGTCAAGGTTTCGTCTCCTGTGTCACCGGATCATCACCCACAAGATGTTTGACCATGTGGTCCTCGTCATCATC
1273  Q  S  R  F  R  L  L  C  H  R  I  I  T  H  K  M  F  D  H  V  V  L  V  I  I

     ************************************
3891 TTCCTCAACTGTATCACCATCGCTATGGAGCGCCCCAAAATTGACCCCCACAGCGCTGAGCGCATCTTCCTGACC
1298  F  L  N  C  I  T  I  A  M  E  R  P  K  I  D  P  H  S  A  E  R  I  F  L  T

     **********************************IIIS2******************************
3966 CTCTCCAACTACATCTTCACGGCAGTCTTTCTAGCTGAAATGACAGTGAAGGTGGTGGCACTGGGCTGGTGCTTT
1323  L  S  N  Y  I  F  T  A  V  F  L  A  E  M  T  V  K  V  V  A  L  G  W  C  F

    ************************************IIIS3**************************************
4041 GGGGAGCAGGCCTACCTGCGCAGCAGCTGGAATGTGCTGGACGGCTTGCTGGTGCTCATCTCCGTCATCGACATC
1348  G  E  Q  A  Y  L  R  S  S  W  N  V  L  D  G  L  L  V  L  I  S  V  I  D  I

     *********                                  *********************************
4116 CTGGTCTCCATGGTCTCCGACAGCGGCACCAAGATCCTTGGCATGCTGAGGGTGCTGCGGCTGCTGCGGACCCTG
1373  L  V  S  M  V  S  D  S  G  T  K  I  L  G  M  L  R  V  L  R  L  L  R  T  L

     ***************IIIS4**********************                          ***
4191 CGTCCACTCAGGGTCATCAGCCGGGCCCAGGGACTGAAGCTGGTGGTAGAGACTCTGATGTCATCCCTCAAACCC
1398  R  P  L  R  V  I  S  R  A  Q  G  L  K  L  V  V  E  T  L  M  S  S  L  K  P

     ************************************IIIS5**********************************
4266 ATTGGCAACATTGTGGTCATTTGCTGTGCCTTCTTCATCATTTTTGGAATTCTCGGGGTGCAGCTCTTCAAAGGG
1423  I  G  N  I  V  V  I  C  C  A  F  F  I  I  F  G  I  L  G  V  Q  L  F  K  G

     ***
4341 AAGTTCTTCGTGTGTCAGGGTGAGGACACCAGGAACATCACTAACAAATCCGACTGCGCTGAGGCCAGCTACCGA
1448  K  F  F  V  C  Q  G  E  D  T  R  N  I  T  N  K  S  D  C  A  E  A  S  Y  R

                                 **********III P Loop**********
4416 TGGGTCCGGCACAAGTACAACTTTGACAACCTGGGCCAGGCTCTGATGTCCCTGTTTGTGCTGGCCTCCAAGGAT
1473  W  V  R  H  K  Y  N  F  D  N  L  G  Q  A  L  M  S  L  F  V  L  A  S  K  D
```

Figure 1D

```
     **********                                                    ***
4491 GGTTGGGTTGACATCATGTATGATGGGCTGGATGCTGTGGGTGTGGATCAGCAGCCCATCATGAACCACAACCCC
1498  G  W  V  D  I  M  Y  D  G  L  D  A  V  G  V  D  Q  Q  P  I  M  N  H  N  P

     **********************************IIIS6************************************
4566 TGGATGCTGCTATACTTCATCTCCTTCCTCCTCATCGTGGCCTTCTTTGTCCTGAACATGTTTGTGGGCGTGGTG
1523  W  M  L  L  Y  F  I  S  F  L  L  I  V  A  F  F  V  L  N  M  F  V  G  V  V

     ************
4641 GTGGAGAACTTCCATAAGTGCAGACAGCACCAGGAGGAGGAGGAGGCGAGGCGGCGTGAGGAGAAGCGACTACGG
1548  V  E  N  F  H  K  C  R  Q  H  Q  E  E  E  E  A  R  R  R  E  E  K  R  L  R

4716 AGGCTGGAGAAAAAGAGAAGGAGTAAGGAGAAGCAGATGGCCGAAGCCCAGTGCAAGCCCTACTACTCTGACTAC
1573  R  L  E  K  K  R  R  S  K  E  K  Q  M  A  E  A  Q  C  K  P  Y  Y  S  D  Y

                                                 *************************IVS1****
4791 TCGAGATTCCGGCTCCTTGTCCACCACCTGTGTACCAGCCACTACCTGGACCTCTTCATCACTGGTGTCATCGGG
1598  S  R  F  R  L  L  V  H  H  L  C  T  S  H  Y  L  D  L  F  I  T  G  V  I  G

     **********************************                          ******
4866 CTGAACGTGGTCACTATGGCCATGGAACATTACCAGCAGCCCCAGATCCTGGACGAGGCTCTGAAGATCTGCAAT
1623  L  N  V  V  T  M  A  M  E  H  Y  Q  Q  P  Q  I  L  D  E  A  L  K  I  C  N

     ***************************IIIS2***************************
4941 TACATCTTTACCGTCATCTTTGTCTTTGAGTCAGTTTTCAAACTTGTGGCCTTTGCGTTCCGCCGTTTCTTCCAG
1648  Y  I  F  T  V  I  F  V  F  E  S  V  F  K  L  V  A  F  A  F  R  R  F  F  Q

       ****************************IVS3*****************************
5016 GACAGGTGGAACCAGCTGGACCTGGCTATTGTGCTTCTGTCCATCATGGGCATCACACTGGAGGAGATTGAGGTC
1673  D  R  W  N  Q  L  D  L  A  I  V  L  L  S  I  M  G  I  T  L  E  E  I  E  V

                                                 *************************IVS4*********
5091 AATCTGTCGCTGCCCATCAACCCCACCATCATCCGTATCATGAGGGTGCTCCGCATTGCTCGAGTTCTGAAGCTG
1698  N  L  S  L  P  I  N  P  T  I  I  R  I  M  R  V  L  R  I  A  R  V  L  K  L

     ****************************                          ******************
5166 TTGAAGATGGCTGTGGGCATGCGGGCACTGCTGCACACGGTGATGCAGGCCCTGCCCCAGGTGGGGAACCTGGGA
1723  L  K  M  A  V  G  M  R  A  L  L  H  T  V  M  Q  A  L  P  Q  V  G  N  L  G

     *************************IVS5*********************************
5241 CTTCTCTTCATGTTATTGTTTTTCATCTTTGCAGGTCTGGGCGTGGAGCTCTTTGGAGACCTGGAGTGTGATGAG
1748  L  L  F  M  L  L  F  F  I  F  A  A  L  G  V  E  L  F  G  D  L  E  C  D  E

                                                 ****************** IV P Loop *******
5316 ACACACCCTTGTGAGGGCTTGGGTCGGCATGCCACCTTTAGGAACTTTGGTATGGCCTTTCTGACCCTCTTCCGA
1773  T  H  P  C  E  G  L  G  R  H  A  T  F  R  N  F  G  M  A  F  L  T  L  F  R

     ************************************
5391 GTCTCCACTGGTGACAACTGGAATGGTATTATGAAGGACCCTTCCCGGGACTGTGACCAGGAGTCCACCTGCTAC
1798  V  S  T  G  D  N  W  N  G  I  M  K  D  P  S  R  D  C  D  Q  E  S  T  C  Y

          ***********************************IVS6*****************************
5466 AACACTGTCATCTCCCCTATCTACTTTGTGTCCTTCGTGCTGACGGCCCAGTTTGTGCTGGTCAACGTGGTCATA
1823  N  T  V  I  S  P  I  Y  F  V  S  F  V  L  T  A  Q  F  V  L  V  N  V  V  I
```

Figure 1E

```
          **********
5541 GCTGTGCTGATGAAGCACCTGGAAGAAAGCAACAAAGAGGCCAAGGAGGAGGCCGAGCTCGAGGCCGAGCTGGAG
1848  A  V  L  M  K  H  L  E  E  S  N  K  E  A  K  E  E  A  E  L  E  A  E  L  E

5616 CTGGAGATGAAGACGCTCAGCCCGCAGCCCCACTCCCCGCTGGGCAGCCCCTTCCTCTGGCCCGGGGTGGAGGGT
1873  L  E  M  K  T  L  S  P  Q  P  H  S  P  L  G  S  P  F  L  W  P  G  V  E  G

5691 GTCAACAGTACTGACAGCCCTAAGCCTGGGGCTCCACACACCACTGCCCACATTGGAGCAGCCTCGGGCTTCTCC
1898  V  N  S  T  D  S  P  K  P  G  A  P  H  T  T  A  H  I  G  A  A  S  G  F  S

5766 CTTGAGCACCCCACGATGGTACCCCACCCCGAGGAGGTGCCAGTCCCCCTAGGACCAGACCTGCTGACTGTGAGG
1923  L  E  H  P  T  M  V  P  H  P  E  E  V  P  V  P  L  G  P  D  L  L  T  V  R

5841 AAGTCTGGTGTCAGCCGGACGCACTCTCTGCCCAATGACAGCTACATGTGCCGCAATGGGAGCACTGCTGAGAGA
1948  K  S  G  V  S  R  T  H  S  L  P  N  D  S  Y  M  C  R  N  G  S  T  A  E  R

5916 TCCCTAGGACACAGGGGCTGGGGGCTCCCCAAAGCCCAGTCAGGCTCCATCTTGTCCGTTCACTCCCAACCAGCA
1973  S  L  G  H  R  G  W  G  L  P  K  A  Q  S  G  S  I  L  S  V  H  S  Q  P  A

5991 GACACCAGCTGCATCCTACAGCTTCCCAAAGATGTGCACTATCTGCTCCAGCCTCATGGGGCTCCCACCTGGGGC
1998  D  T  S  C  I  L  Q  L  P  K  D  V  H  Y  L  L  Q  P  H  G  A  P  T  W  G

6066 GCCATCCCTAAACTACCCCCACCTGGCCGCTCCCCTCTGGCTCAGAGGCCTCTCAGGCGCCAGGCAGCAATAAGG
2023  A  I  P  K  L  P  P  P  G  R  S  P  L  A  Q  R  P  L  R  R  Q  A  A  I  R

6141 ACTGACTCCCTGGATGTGCAGGGCCTGGGTAGCCGGGAAGACCTGTTGTCAGAGGTGAGTGGGCCCTCCTGCCCT
2048  T  D  S  L  D  V  Q  G  L  G  S  R  E  D  L  L  S  E  V  S  G  P  S  C  P

6216 CTGACCCGGTCCTCATCCTTCTGGGGCGGGTCGAGCATCCAGGTGCAGCAGCGTTCCGGCATCCAGAGCAAAGTC
2073  L  T  R  S  S  S  F  W  G  G  S  S  I  Q  V  Q  Q  R  S  G  I  Q  S  K  V

6291 TCCAAGCACATCCGCCTGCCAGCCCCTTGCCCAGGCCTGGAACCCAGCTGGGCCAAGGACCCTCCAGAGACCAGA
2098  S  K  H  I  R  L  P  A  P  C  P  G  L  E  P  S  W  A  K  D  P  P  E  T  R

6366 AGCAGCTTAGAGCTGGACACGGAGCTGAGCTGGATTTCAGGAGACCTCCTTCCCAGCAGCCAGGAAGAACCCCTG
2123  S  S  L  E  L  D  T  E  L  S  W  I  S  G  D  L  L  P  S  S  Q  E  E  P  L

6441 TTCCCACGGGACCTGAAGAAGTGCTACAGTGTAGAGACCCAGAGCTGCAGGCGCAGGCCTGGGTTCTGGCTAGAT
2148  F  P  R  D  L  K  K  C  Y  S  V  E  T  Q  S  C  R  R  R  P  G  F  W  L  D

6516 GAACAGCGGAGACACTCCATTGCTGTCAGCTGTCTGGACAGCGGCTCCCAACCCCGCCTATGTCCAAGCCCCTCA
2173  E  Q  R  R  H  S  I  A  V  S  C  L  D  S  G  S  Q  P  R  L  C  P  S  P  S

6591 AGCCTCGGGGGCCAACCTCTTGGGGGTCCTGGGAGCCGGCCTAAGAAAAAACTCAGCCCACCCAGTATCTCTATA
2198  S  L  G  G  Q  P  L  G  G  P  G  S  R  P  K  K  K  L  S  P  P  S  I  S  I

6666 GACCCCCCGGAGAGCCAGGGCTCTCGGCCCCCATGCAGTCCTGGTGTCTGCCTCAGGAGGAGGGCGCCGGCCAGT
2223  D  P  P  E  S  Q  G  S  R  P  P  C  S  P  G  V  C  L  R  R  R  A  P  A  S

6741 GACTCTAAGGATCCCTCGGTCTCCAGCCCCCTTGACAGCACGGCTGCCTCACCCTCCCCAAAGAAAGACACGCTG
2248  D  S  K  D  P  S  V  S  S  P  L  D  S  T  A  A  S  P  S  P  K  K  D  T  L

6816 AGTCTCTCTGGTTTGTCTTCTGACCCAACAGACATGGACCCCTG  SEQ ID NO:1
2273  S  L  S  G  L  S  S  D  P  T  D  M  D  P  @  SEQ ID NO:1
```

Figure 1F

```
  1 ATGACCGAGGGCGCACGGGCCGCCGACGAGGTCCGGGTGCCCCTGGGGCGCCGCCCCTGGCCCTGCGGCGTTGGT
 26  M  T  E  G  A  R  A  A  D  E  V  R  V  P  L  G  R  R  P  W  P  C  G  V  G

 76 GGGGGCGTCCCCGGAGAGCCCCGGGGCGCCGGGACGCGAGGCGGAGGGGGGTTCGAGCTCGGCGTGTCACCCTCC
 51  G  G  V  P  G  E  P  R  G  A  G  T  R  G  G  G  G  F  E  L  G  V  S  P  S

151 GAGAGCCCGGCGGCCGAGCGCTGCGCGGAGCTGGGTGCCGACGAGGAGCAGCGCGTCCCGTACCCGGCCTTGGCG
 76  E  S  P  A  A  E  R  C  A  E  L  G  A  D  E  E  Q  R  V  P  Y  P  A  L  A
                                                                      ******

226 GCCACGGTCTTCTTCTGCCTCGGTCAGACCACGCGGCCGCGCAGCTGGTCCGTCCGGCTGGTCTGCAACCCATGG
101  A  T  V  F  F  C  L  G  Q  T  T  R  P  R  S  W  S  V  R  L  V  C  N  P  W

    *************************IS1************************************

301 TTCGAGCACGTGAGCATGCTGGTAATCATGCTCAACTGCGTGACCCTGGGCATGTTCCGGCCCTGTGAGGACGTT
126  F  E  H  V  S  M  L  V  I  M  L  N  C  V  T  L  G  M  F  R  P  C  E  D  V

                                            *************************IS2***********

376 GAGTGCGGCTCCGAGCGCTGCAACATCCTGGAGGCCTTTGACGCCTTCATTTTCGCCTTTTTTGCGGTGGAGATG
151  E  C  G  S  E  R  C  N  I  L  E  A  F  D  A  F  I  F  A  F  F  A  V  E  M

    ****************************                  **********************

451 GTCATCAAGATGGTGGCCTTGGGGCTGTTCGGGCAGAAGTGTTACCTGGGTGACACGTGGAACAGGCTGGATTTC
176  V  I  K  M  V  A  L  G  L  F  G  Q  K  C  Y  L  G  D  T  W  N  R  L  D  F

    **************IS3**********************                  ******************

526 TTCATCGTCGTGGCGGGCATGATGGAGTACTCGTTGGACGGACACAACGTGAGCCTCTCGGCTATCAGGACCGTG
201  F  I  V  V  A  G  M  M  E  Y  S  L  D  G  H  N  V  S  L  S  A  I  R  T  V


    ************IS4************************************

601 CGGGTGCTGCGGCCCCTCCGCGCCATCAACCGCGTGCCTAGCATGCGGATCCTGGTCACTCTGCTGCTGGATACG
226  R  V  L  R  P  L  R  A  I  N  R  V  P  S  M  R  I  L  V  T  L  L  L  D  T


          ********************************IS5************************************

676 CTGCCCATGCTCGGGAACGTCCTTCTGCTGTGCTTCTTCGTCTTCTTCATTTTCGGCATCGTTGGCGTCCAGCTC
251  L  P  M  L  G  N  V  L  L  L  C  F  F  V  F  F  I  F  G  I  V  G  V  Q  L


    ****
751 TGGGCTGGCCTCCTGCGGAACCGCTGCTTCCTGGACAGTGCCTTTGTCAGGAACAACAACCTGACCTTCCTGCGG
276  W  A  G  L  L  R  N  R  C  F  L  D  S  A  F  V  R  N  N  N  L  T  F  L  R


826 CCGTACTACCAGACGGAGGAGGGCGAGGAGAACCCGTTCATCTGCTCCTCACGCCGAGACAACGGCATGCAGAAG
301  P  Y  Y  Q  T  E  E  G  E  E  N  P  F  I  C  S  S  R  R  D  N  G  M  Q  K


901 TGCTCGCACATCCCCGGCCGCCGCGACGTGCGCATGCCCTGCACCCTGGGCTGGGAGGCCTACACGCAGCCGCAG
326  C  S  H  I  P  G  R  R  D  V  R  M  P  C  T  L  G  W  E  A  Y  T  Q  P  Q


976 GCCGAGGGGGTGGGCGCTGCACGCAACGCCTGCATCAACTGGAACCAGTACTACAACGTGTGCCGCTCGGGTGAC
351  A  E  G  V  G  A  A  R  N  A  C  I  N  W  N  Q  Y  Y  N  V  C  R  S  G  D
```

Figure 2A

```
                                ****************************I P
Loop***************
1051 TCCAACCCCCACAACGGTGCCATCAACTTCGACAACACCTGCTACGCCTGGATTGCCATCTTCCAGGTGATCACG
 376  S  N  P  H  N  G  A  I  N  F  D  N  T  C  Y  A  W  I  A  I  F  Q  V  I  T

     ****************************                    ****************************
1126 CTGGAAGGCTGGGTGGACATCATGTACTACGTCATGGACGCCCACTCATTCTACAACTTCATCTATTTCATCCTG
 401  L  E  G  W  V  D  I  M  Y  Y  V  M  D  A  H  S  F  Y  N  F  I  Y  F  I  L

     **********************IS6************************************
1201 CTCATCATCGTGGGCTCCTTCTTCATGATCAACCTGTGCCTGGTGGTGATTGCCACGCAGTTCTCGGAGACGAAG
 426  L  I  I  V  G  S  F  F  M  I  N  L  C  L  V  V  I  A  T  Q  F  S  E  T  K

1276 CAGCGGGAGAGTCAGCTGATGCGGGAGCAGCGGGCACGCCACCTGTCCAACGACAGCACGCTGGCCAGCTTCTCC
 451  Q  R  E  S  Q  L  M  R  E  Q  R  A  R  H  L  S  N  D  S  T  L  A  S  F  S

1351 GAGCCTGGCAGCTGCTACGAAGAGCTGCTGAAGTACGTGGGCCACATATTCCGCAAGGTCAAGCGGCAGCTTGCG
 476  E  P  G  S  C  Y  E  E  L  L  K  Y  V  G  H  I  F  R  K  V  K  R  Q  L  A

1426 CCTCTACGCCCGCTGGCAGAGCCGTGGCGCAAGAAGGTGGACCCCAGTGCTGTGCAAGGCCAGGGTCCCGGGCAC
 501  P  L  R  P  L  A  E  P  W  R  K  K  V  D  P  S  A  V  Q  G  Q  G  P  G  H

1501 CGCCAGCGCCGGGCAGGCAGGCACACAGCCTCGGTGCACCACCTGGTCTACCACCACCATCACCACCACCACCAC
 526  R  Q  R  R  A  G  R  H  T  A  S  V  H  H  L  V  Y  H  H  H  H  H  H  H  H

1576 CACTACCATTTCAGCCATGGCAGCCCCCGCAGGCCCGGCCCCGAGCCAGGCGCCTGCGACACCAGGCTGGTCCGA
 551  H  Y  H  F  S  H  G  S  P  R  R  P  G  P  E  P  G  A  C  D  T  R  L  V  R

1651 GCTGGCGCGCCCCCCTCGCCACCTTCCCCAGGCCGCGGACCCCCCGACGCAGAGTCTGTGCACAGCATCTACCAT
 576  A  G  A  P  P  S  P  P  S  P  G  R  G  P  P  D  A  E  S  V  H  S  I  Y  H

1726 GCCGACTGCCACATAGAGGGGCCGCAGGAGAGGGCCCGGGTGGGCACATGCCGCAGCCACTGCCGCTGCCAGCCT
 601  A  D  C  H  I  E  G  P  Q  E  R  A  R  V  G  T  C  R  S  H  C  R  C  Q  P

1801 CAGGCTGGCCACAGGGCTGGGCACCATGAACTACCCCACGATCCTGCCCTCAGGGGTGGGCAGCGGCAAAGGCAG
 626  Q  A  G  H  R  A  G  H  H  E  L  P  H  D  P  A  L  R  G  G  Q  R  Q  R  Q

1876 CACCAGCCCCGGACCCAAGGGGAAGTGGGCCGGTGGACCGCCAGGCACCGGGGGGCACGGCCCGTTGAGCTTGAAC
 651  H  Q  P  R  T  Q  G  E  V  G  R  W  T  A  R  H  R  G  H  G  P  L  S  L  N

1951 AGCCCTGATCCCTACGAGAAGATCCCGCATGTGGCCGGGGAGCATGGACTGGCCAGCCCTGGCCATCTGTCGGGC
 676  S  P  D  P  Y  E  K  I  P  H  V  A  G  E  H  G  L  A  S  P  G  H  L  S  G
```

Figure 2B

```
2026 CTCAGTGTGCCCTGCCCCCTGCCCAGCCCCCCAGCGGGCACACTGACCTGTGAGCTGAAGAGCTGCCCGTACTGC
 701  L  S  V  P  C  P  L  P  S  P  P  A  G  T  L  T  C  E  L  K  S  C  P  Y  C

2101 ACCCGTGCCCTGGAGGACCCGGAGGGTGAGCTCAGCGGCTCGGAAAGTGGAGACTCAGATGGCCGTGGCGTCTAT
 726  T  R  A  L  E  D  P  E  G  E  L  S  G  S  E  S  G  D  S  D  G  R  G  V  Y

2176 GAATTCACGCAGGACGTCCGGCACGGTGACCGCTGGGACCCCACGCGACCACCCCGTGCGACGGACACACCAGGC
 751  E  F  T  Q  D  V  R  H  G  D  R  W  D  P  T  R  P  P  R  A  T  D  T  P  G

2251 CCAGGCCCAGGCAGCCCCCAGCGGCGGGCACAGCAGAGGGCAGCCCCGGGCGAGCCAGGCTGGATGGGCCGCCTC
 776  P  G  P  G  S  P  Q  R  R  A  Q  Q  R  A  A  P  G  E  P  G  W  M  G  R  L

                                                  **********IIS1*****
2326 TGGGTTACCTTCAGCGGCAAGCTGCGCCGCATCGTGGACAGCAAGTACTTCAGCCGTGGCATCATGATGGCCATC
 801  W  V  T  F  S  G  K  L  R  R  I  V  D  S  K  Y  F  S  R  G  I  M  M  A  I

     ************************************                          ***
2401 CTTGTCAACACGCTGAGCATGGGCGTGGAGTACCATGAGCAGCCCGAGGAGCTGACTAATGCTCTGGAGATCAGC
 826  L  V  N  T  L  S  M  G  V  E  Y  H  E  Q  P  E  E  L  T  N  A  L  E  I  S

     **********IIS2**************************************
2476 AACATCGTGTTCACCAGCATGTTTGCCCTGGAGATGCTGCTGAAGCTGCTGCGCGCTGTCCCTCTGGGCTACATC
 851  N  I  V  F  T  S  M  F  A  L  E  M  L  L  K  L  L  R  A  V  P  L  G  Y  I

                **********IIS3**********
2551 CGGAACCCGTACAACATCTTCGACGGCATCATCGTGGTCATCAGCGTCTGGGAGATCGTGGGGCAGGCGGACGGT
 876  R  N  P  Y  N  I  F  D  G  I  I  V  V  I  S  V  W  E  I  V  G  Q  A  D  G

                **********IIS4**********
2626 GGCTTGTCTGTGCTGCGCACCTTCCGGCTGCTGCGTGTGCTGAAGCTGGTGCGCTTTCTGCCAGCCCTGCGGCGC
 901  G  L  S  V  L  R  T  F  R  L  L  R  V  L  K  L  V  R  F  L  P  A  L  R  R

     ******                          **********IIS5**********
2701 CAGCTCGTGGTGCTGGTGAAGACCATGGACAACGTGGCTACCTTCTGCACGCTGCTCATGCTCTTCATTTTCATC
 926  Q  L  V  V  L  V  K  T  M  D  N  V  A  T  F  C  T  L  L  M  L  F  I  F  I

     ************************************
2776 TTCAGCATCCTGGGCATGCACCTTTTCGGCTGCAAGTTCAGCCTGAAGACAGACACCGGAGACACCGTGCCTGAC
 951  F  S  I  L  G  M  H  L  F  G  C  K  F  S  L  K  T  D  T  G  D  T  V  P  D

                **********II P Loop**********
2851 AGGAAGAACTTCGACTCCCTGCTGTGGGCCATCGTCACCGTGTTCCAGATCCTGACCCAGGAGGACTGGAACGTG
 976  R  K  N  F  D  S  L  L  W  A  I  V  T  V  F  Q  I  L  T  Q  E  D  W  N  V
```

Figure 2C

```
     *********                                 **********************************IIS6*********
2926 GTCCTGTACAACGGCATGGCCTCCACCTCCTCCTGGGCCGCCCTCTACTTCGTGGCCCTCATGACCTTCGGCAAC
1001  V  L  Y  N  G  M  A  S  T  S  S  W  A  A  L  Y  F  V  A  L  M  T  F  G  N

     **********************************************
3001 TATGTGCTCTTCAACCTGCTGGTGGCCATCCTCGTGGAGGGCTTCCAGGCGGAGGGCGATGCCAACAGATCCGAC
1026  Y  V  L  F  N  L  L  V  A  I  L  V  E  G  F  Q  A  E  G  D  A  N  R  S  D

3076 ACGGACGAGGACAAGACGTCGGTCCACTTCGAGGAGGACTTCCACAAGCTCAGAGAACTCCAGACCACAGAGCTG
1051  T  D  E  D  K  T  S  V  H  F  E  E  D  F  H  K  L  R  E  L  Q  T  T  E  L

3151 AAGATGTGTTCCCTGGCCGTGACCCCCAACGGCACCTGGAGGGACGAGGCAGCCTGTCCCCTCCCCTCATCATGT
1076  K  M  C  S  L  A  V  T  P  N  G  T  W  R  D  E  A  A  C  P  L  P  S  S  C

3226 GCACAGCTGCCACGCCCATGCCTACCCCCAAGAGCTCACCATTCCTGGATGCAGCCCCCAGCCTCCCAGACTCTC
1101  A  Q  L  P  R  P  C  L  P  P  R  A  H  H  S  W  M  Q  P  P  A  S  Q  T  L

3301 GGCGTGGCAGCAGCAGCTCCGGGGACCCGCCACTGGGAGACCAGAAGCCTCCGGCAGCCTCCGAAGTTCTCCCTG
1126  G  V  A  A  A  A  P  G  T  R  H  W  E  T  R  S  L  R  Q  P  P  K  F  S  L

3376 TGCCCCCTGGGGCCCAGTGGCGCCTGGAGCAGCCGGCGCTCCAGCTGGAGCAGCCTGGGCCGTGCCCAGCCTCAA
1151  C  P  L  G  P  S  G  A  W  S  S  R  R  S  S  W  S  S  L  G  R  A  Q  P  Q

3451 GCGCCGGCGTGCCAGTGTGGGGAACGTGAGTCCCTGCTGTCTGGCGAGGGCAAGGGCAGCACCGACGACGAAGCT
1176  A  P  A  C  Q  C  G  E  R  E  S  L  L  S  G  E  G  K  G  S  T  D  D  E  A

3526 GAGGACGGCAGGGCGCGCTCCGGGCCCCGTGCCACCCCACTGCGGCGGGCCGAGTCCCTGGACCCACGGCCCCTG
1201  E  D  G  R  A  R  S  G  P  R  A  T  P  L  R  R  A  E  S  L  D  P  R  P  L

3601 CGGCGGCCGCCTCCCGCCTACCAAGTGCGCGATCGCGACGGGCAGGTGGTGGCCCTGCCCAGCGACTTCTTCCTG
1226  R  R  P  P  P  A  Y  Q  V  R  D  R  D  G  Q  V  V  A  L  P  S  D  F  F  L

3676 CGCATCGACAGCCACCGTGAGGATGCAGCCGAGCTTGACGACGACTCGGAGGACAGCTGCTGCCTCCGCCTGCAT
1251  R  I  D  S  H  R  E  D  A  A  E  L  D  D  D  S  E  D  S  C  C  L  R  L  H

3751 AAAGTGCTGGTGCCCTACAAGCCCCAGCGGTGCCGGAGCAGGAGGCCTGGGCCCTCTACCCTCTACCTCTTCTCC
1276  K  V  L  V  P  Y  K  P  Q  R  C  R  S  R  R  P  G  P  S  T  L  Y  L  F  S

                                                  ******************IIIS1*****
3826 CCACAGAACCGGTTCCGCGTCTCCTGCCAGAAGGTCATCACACACAAGATGTTTGATCACGTGGTCCTCGTCTTC
1301  P  Q  N  R  F  R  V  S  C  Q  K  V  I  T  H  K  M  F  D  H  V  V  L  V  F

     ****************************************
3901 ATCTTCCTCAACTGCGTCACCATCGCCCTGGAGAGGCCTGACATTGATCCCGGCAGCACCGAGCGGGTCTTCCTC
1326  I  F  L  N  C  V  T  I  A  L  E  R  P  D  I  D  P  G  S  T  E  R  V  F  L

     ********************************IIIS2************************************
3976 AGCGTCTCCAATTACATCTTCACGGCCATCTTCGTGGCGGAGATGATGGTGAAGGTGGTGGCCCTGGGGCTGCTG
1351  S  V  S  N  Y  I  F  T  A  I  F  V  A  E  M  M  V  K  V  V  A  L  G  L  L

                                        ************************IIIS3*****************
4051 TCCGGCGAGCACGCCTACCTGCAGAGCAGCTGGAACCTGCTGGATGGGCTGCTGGTGCTGGTGTCCCTGGTGGAC
1376  S  G  E  H  A  Y  L  Q  S  S  W  N  L  L  D  G  L  L  V  L  V  S  L  V  D
```

Figure 2D

```
                                                          ****************************
4126 ATTGTCGTGGCCATGGCCTCGGCTGGTGGCGCCAAGATCCTGGGTGTTCTGCGCGTGCTGCGTCTGCTGCGGACC
1401  I  V  V  A  M  A  S  A  G  G  A  K  I  L  G  V  L  R  V  L  R  L  L  R  T

     *********IIIS4**************************                                  ***
4201 CTGCGGCCTCTGAGGGTCATCAGCCGGCCCCGGCTCAAGCTGGTGGTGGAGACGCTGATATCATCACTCAGGCCC
1426  L  R  P  L  R  V  I  S  R  P  R  L  K  L  V  V  E  T  L  I  S  S  L  R  P

     ***********************************IIIS5*************************************
4276 ATTGGGAACATCGTCCTCATCTGCTGCGCCTTCTTCATCATTTTTGGCATTTTGGGTGTGCAGCTCTTCAAAGGG
1451  I  G  N  I  V  L  I  C  C  A  F  F  I  I  F  G  I  L  G  V  Q  L  F  K  G

     ***
4351 AAGTTCTACTACTGCGAGGGCCCCGACACCAGGAACATCTCCACCAAGGCACAGTGCCGGGCCGCCCACTACCGC
1476  K  F  Y  Y  C  E  G  P  D  T  R  N  I  S  T  K  A  Q  C  R  A  A  H  Y  R

                           **************************III P Loop***************
4426 TGGGTGCGACGCAAGTACAACTTCGACAACCTGGGCCAGGCCCTGATGTCGCTGTTCGTGCTGTCATCCAAGGAT
1501  W  V  R  R  K  Y  N  F  D  N  L  G  Q  A  L  M  S  L  F  V  L  S  S  K  D

     **********************                                                    ***
4501 GGATGGGTGAACATCATGTACGACGGGCTGGATGCCGTGGGTGTCGACCAGCAGCCTGTGCAGAACCACAACCCC
1526  G  W  V  N  I  M  Y  D  G  L  D  A  V  G  V  D  Q  Q  P  V  Q  N  H  N  P

     ****************************************IIIS6*******************************
4576 TGGATGCTGCTGTACTTCATCTCCTTCCTCTGCTACATCGTCAGCTTCTTCGTGCTCAACATGTTCGTGGGCGTC
1551  W  M  L  L  Y  F  I  S  F  L  C  Y  I  V  S  F  F  V  L  N  M  F  V  G  V

     ***************
4651 GTGGTCGAGAACTTCCACAAGTGCCGGCCGCACCAGGAGGCGGAGGAGGCGCGGCGGCGAGAGGAGAAGCGGCTG
1576  V  V  E  N  F  H  K  C  R  P  H  Q  E  A  E  E  A  R  R  R  E  E  K  R  L

4726 CGGCGCCTAGAGAGGAGGCGCAGGAGCACTTTCCCCAGCCCAGAGGCCCAGCGCCGGCCCTACTATGCCGACTAC
1601  R  R  L  E  R  R  R  R  S  T  F  P  S  P  E  A  Q  R  R  P  Y  Y  A  D  Y

                                                  *************************IVS1***
4801 TCGCCCACGCGCCGCCGCTCCATTCACTCGCTGTGCACCAGCCACTATCTCGACCTCTTCATCACCTTCATCATC
1626  S  P  T  R  R  R  S  I  H  S  L  C  T  S  H  Y  L  D  L  F  I  T  F  I  I

     ************************************                                     ***
4876 TGTGTCAACGTCATCACCATGTCCATGGAGCACTATAACCAACCCAAGTCGCTGGACGAGGCCCTCAAGTACTGC
1651  C  V  N  V  I  T  M  S  M  E  H  Y  N  Q  P  K  S  L  D  E  A  L  K  Y  C

     ************************IVS2************************************
4951 AACTACGTCTTCACCATCGTGTTTGTCTTCGAGGCTGCACTGAAGCTGGTAGCATTTGGGTTCCGTCGGTTCTTC
1676  N  Y  V  F  T  I  V  F  V  F  E  A  A  L  K  L  V  A  F  G  F  R  R  F  F

          ************************IVS3*******************************
5026 AAGGACAGGTGGAACCAGCTGGACCTGGCCATCGTGCTGCTGTCACTCATGGGCATCACGCTGGAGGAGATAGAG
1701  K  D  R  W  N  Q  L  D  L  A  I  V  L  L  S  L  M  G  I  T  L  E  E  I  E
```

Figure 2E

```
                                          **************************IVS4**************
5101 ATGAGCGCCGCGCTGCCCATCAACCCCACCATCATCCGCATCATGCGCGTGCTTCGCATTGCCCGTGTGCTGAAG
1726  M  S  A  A  L  P  I  N  P  T  I  I  R  I  M  R  V  L  R  I  A  R  V  L  K

     *********************************                                 ***************
5176 CTGCTGAAGATGGCTACGGGCATGCGCGCCCTGCTGGACACTGTGGTGCAAGCTCTCCCCCAGGTGGGGAACCTG
1751  L  L  K  M  A  T  G  M  R  A  L  L  D  T  V  V  Q  A  L  P  Q  V  G  N  L

     *************************IVS5*************************************
5251 GGCCTTCTTTTCATGCTCCTGTTTTTTATCTATCTGAGATTGGGAGTGGAGCTGTTCGGGAGGCTGGAGTGCAGT
1776  G  L  L  F  M  L  L  F  F  I  Y  L  R  L  G  V  E  L  F  G  R  L  E  C  S

                                                  ****************IV P Loop*******
5326 GAAGACAACCCCTGCGAGGGCCTGAGCAGGCACGCCACCTTCAGCAACTTCGGCATGGCCTTCCTCACGCTGTTC
1801  E  D  N  P  C  E  G  L  S  R  H  A  T  F  S  N  F  G  M  A  F  L  T  L  F

     ****************************************
5401 CGCGTGTCCACGGGGGACAACTGGAACGGGATCATGAAGGACACGCTGCGCGAGTGCTCCCGTGAGGACAAGCAC
1826  R  V  S  T  G  D  N  W  N  G  I  M  K  D  T  L  R  E  C  S  R  E  D  K  H

                            *************************IVS6*****************************
5476 TGCCTGAGCTACCTGCCGGCCCCGTCGCCCGTCTACTTCGTGACCTTCGTGCTGGTGCCCCAGTTCGTGCTGGTG
1851  C  L  S  Y  L  P  A  P  S  P  V  Y  F  V  T  F  V  L  V  P  Q  F  V  L  V

     **********************************
5551 AACGTGGTGGTGGCCGTGCTCATGAAGCACCTGGAGGAGAGCAACAAGGAGGCTCGGGAGGATGCGGAGCTGGAC
1876  N  V  V  V  A  V  L  M  K  H  L  E  E  S  N  K  E  A  R  E  D  A  E  L  D

5626 GCCGAGATCGAGCTGGAGATGGCGCAGGGCCCCGGGAGTGCACGCCGGGTGGACGCGGACAGGCCTCCCTTGCCC
1901  A  E  I  E  L  E  M  A  Q  G  P  G  S  A  R  R  V  D  A  D  R  P  P  L  P

5701 CAGGAGAGTCCGGCGCCAGGGACGCCCCAAACCTGGTTGCACGCAAGGTGTCCGTGTCCAGGATCTCTCGCTGCC
1926  Q  E  S  P  A  P  G  T  P  Q  T  W  L  H  A  R  C  P  C  P  G  S  L  A  A

5776 CAACGACAGCTACATGTTCAGGCCCGTGGTGCCTGCCTCGGCGCCCCGGGCCCGCCCGCTGCAGGAGGTGGAGAT
1951  Q  R  Q  L  H  V  Q  A  R  G  A  C  L  G  A  P  G  P  P  A  A  G  G  G  D

5851 GGAGACCTATGGGGCCGGCACCCCCTTGGAGTCCTGTGCCATCCCATCCAGATCCCATTGGCTGTGTCGAACCCA
1976  G  D  L  W  G  R  H  P  L  G  V  L  C  H  P  I  Q  I  P  L  A  V  S  N  P

5926 GCCAGGAGCGGCGAGCCCCTCCACGCCCTGTCCCCTCGGGGCACAGCCGCTCCCCCAGTCTCAGCCGGCTGCTCT
2001  A  R  S  G  E  P  L  H  A  L  S  P  R  G  T  A  A  P  P  V  S  A  G  C  S

6001 GCAGACAGGAGGCTGTGCACACCGATTCCTTGGAAGGGAAGATTGACAGCCCTAGGGACACCCTGGATCCTGCAG
2026  A  D  R  R  L  C  T  P  I  P  W  K  G  R  L  T  A  L  G  T  P  W  I  L  Q

6076 AGCCTGGTGAGAAACCCCCCGG SEQ ID NO:3
2051  S  L  V  R  N  P  R  SEQ ID NO:3
```

Figure 2F

α1G	RIMRVLRIARVLKLLKMA	SEQ ID NO 7
α1H	RIMRVLRIARVLKLLKMA	SEQ ID NO 7
α1S	RLFRVMRLIKLLSRAEGV	SEQ ID NO 8
α1C	RLFRVMRLVKLLSRGEGI	SEQ ID NO 9
α1D	RLFRVMRLVKLLSRGEGI	SEQ ID NO 9
α1A	RLFRAARLIKLLRQGYTI	SEQ ID NO 10
α1B	RLFRAARLIKLLRQGYTI	SEQ ID NO 10
α1E	KLFRAARLIKLLRQGYTI	SEQ ID NO 10

Fig. 3

Fig. 4A

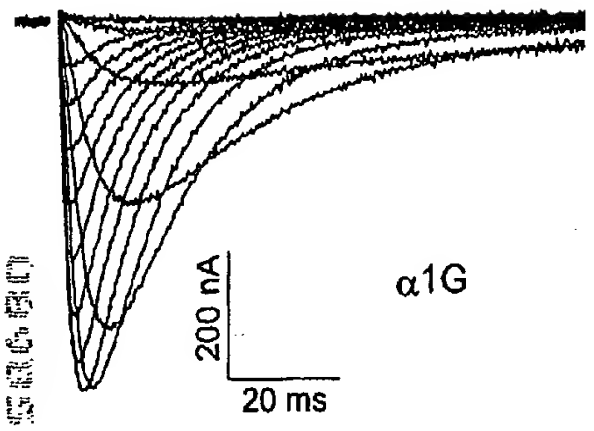

Fig. 4B

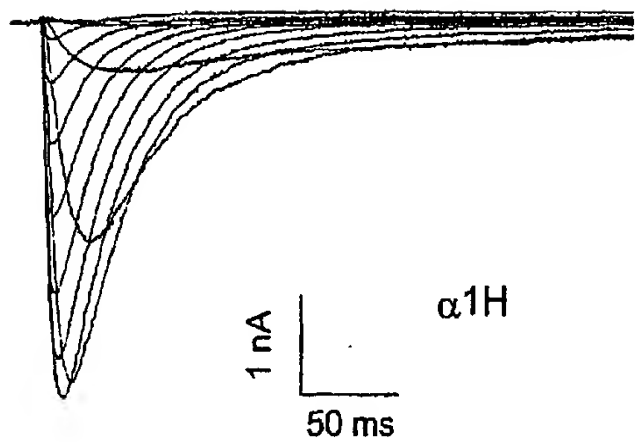

Fig. 4C

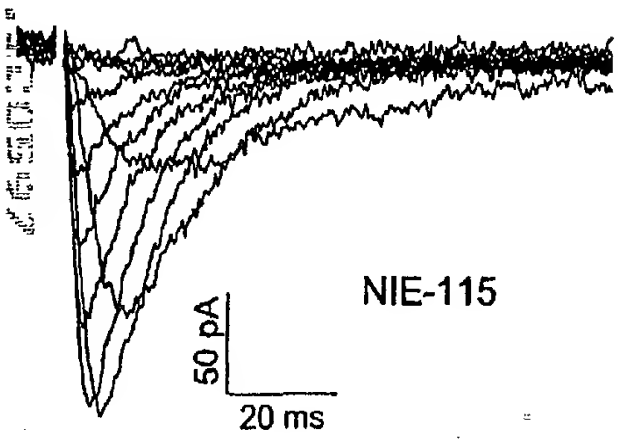

Fig. 4D

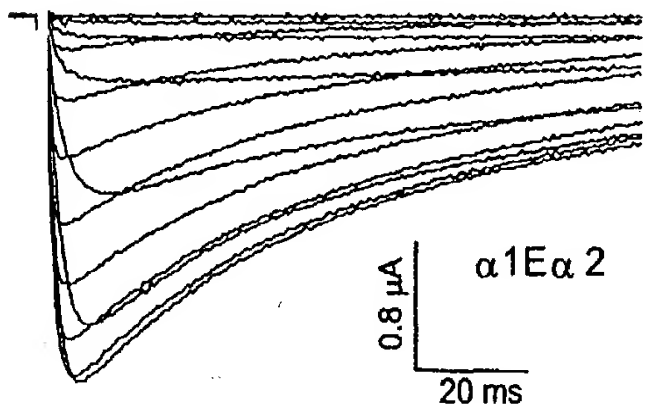

Fig. 5A

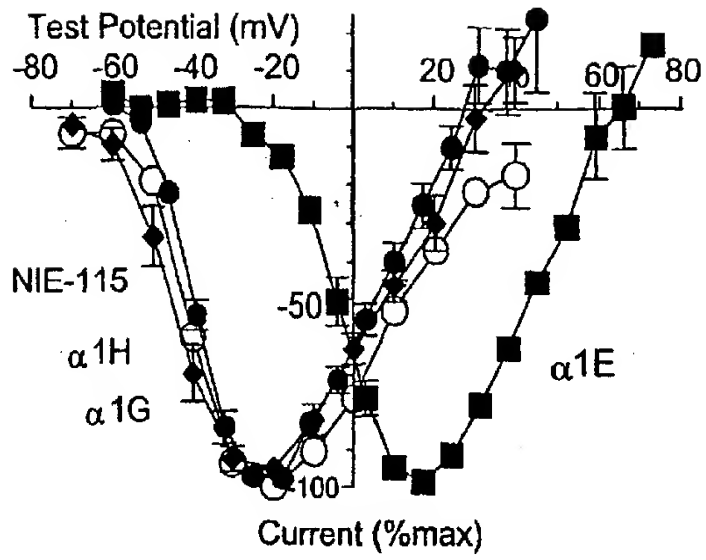

Fig. 5B

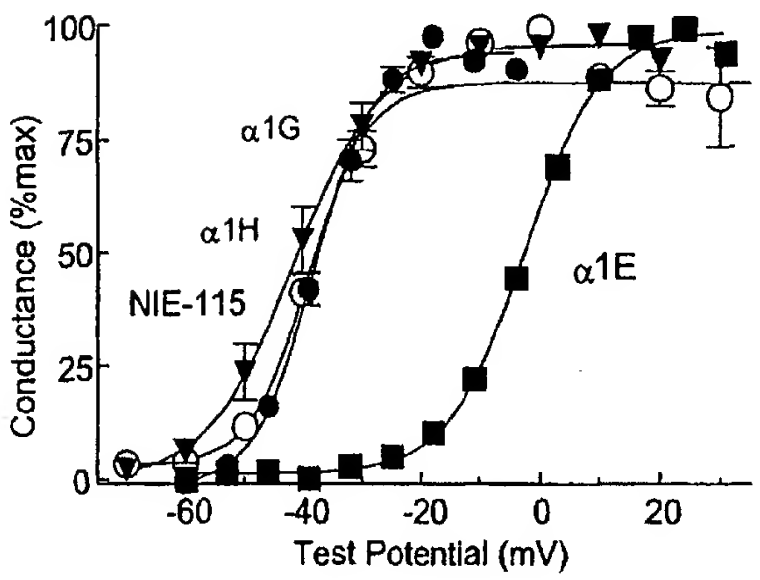

Fig. 5C

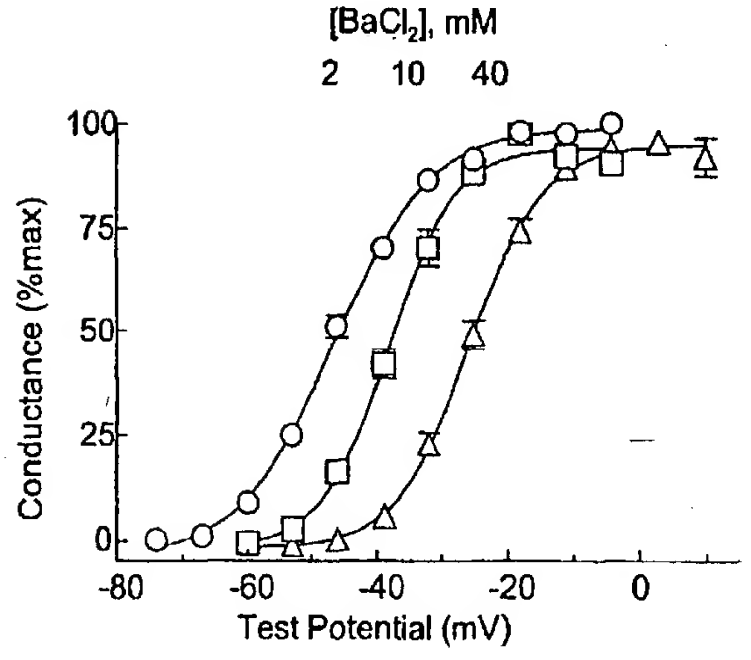

Fig. 6A

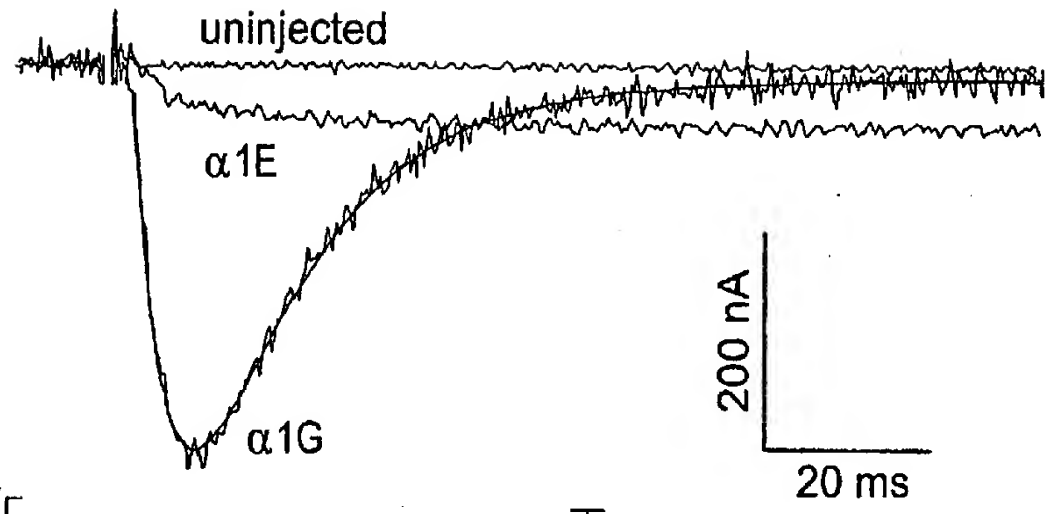
uninjected
α1E
α1G
200 nA
20 ms

Fig. 6B

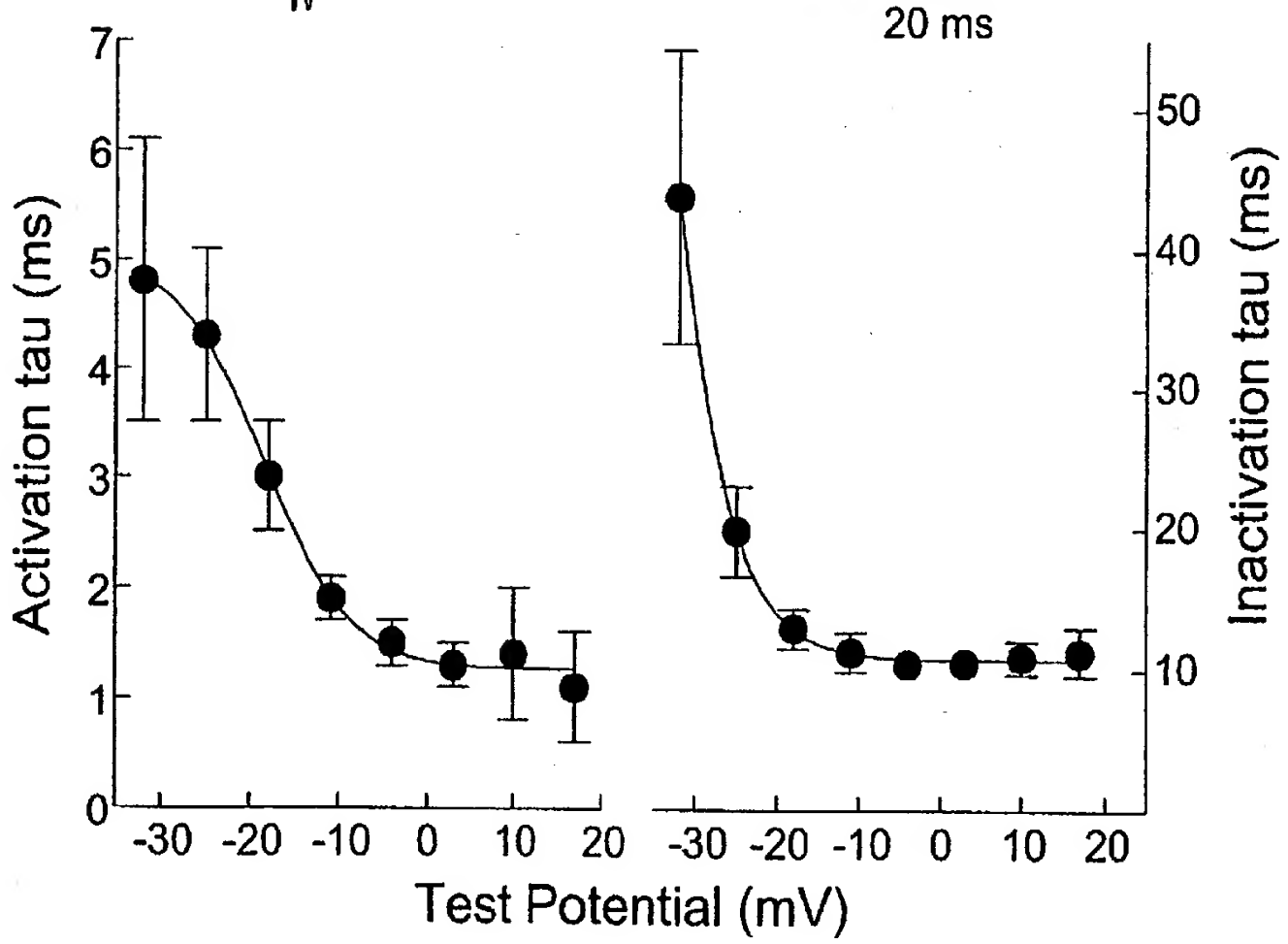
Activation tau (ms)
Inactivation tau (ms)
Test Potential (mV)
0
1
2
3
4
5
6
7
10
20
30
40
50
-30
-20
-10
0
10
20

Fig. 7A

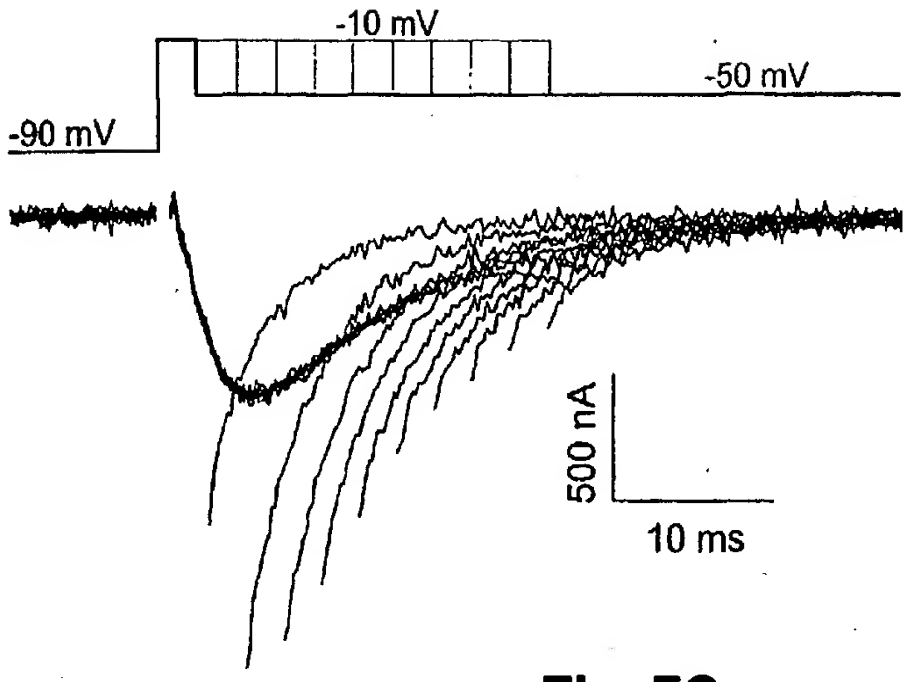

Fig. 7B

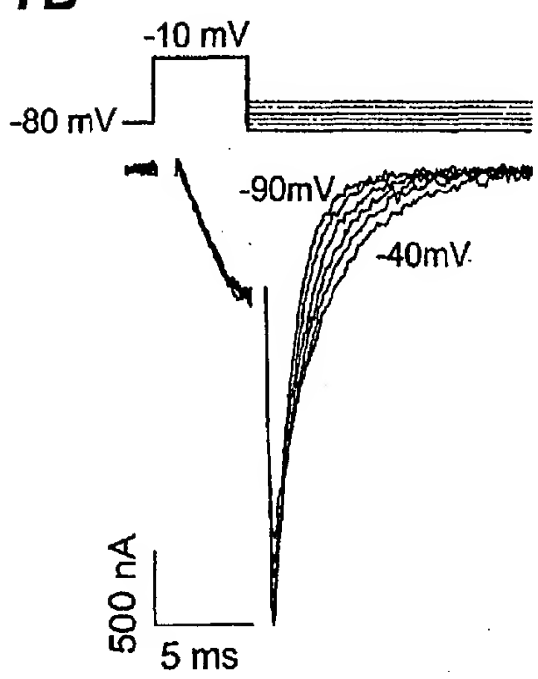

Fig. 7C

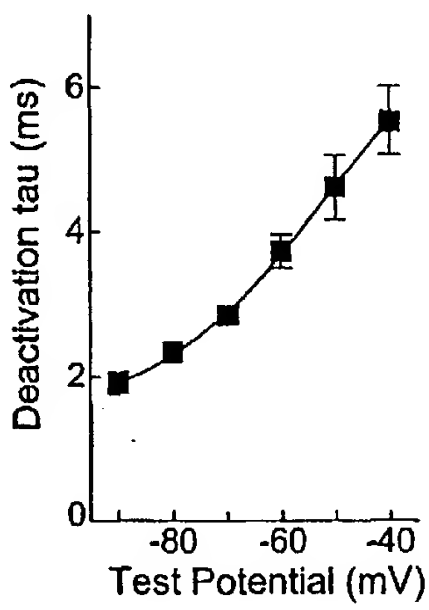

Fig. 7d

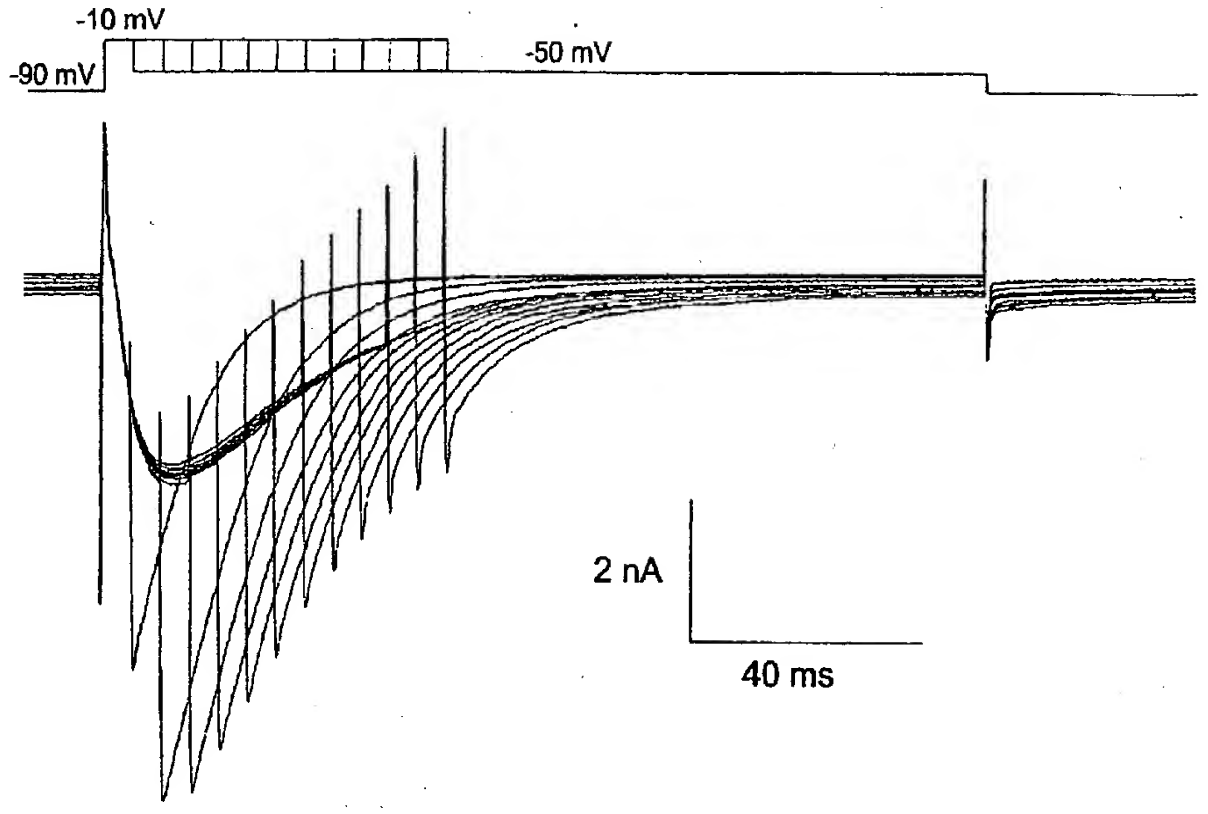

Fig. 7e

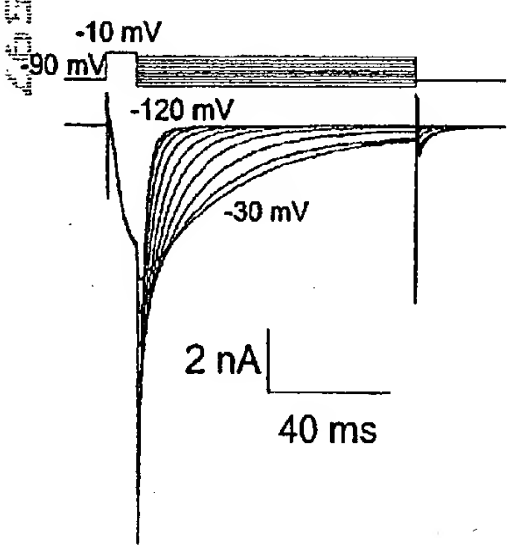

Fig. 7f

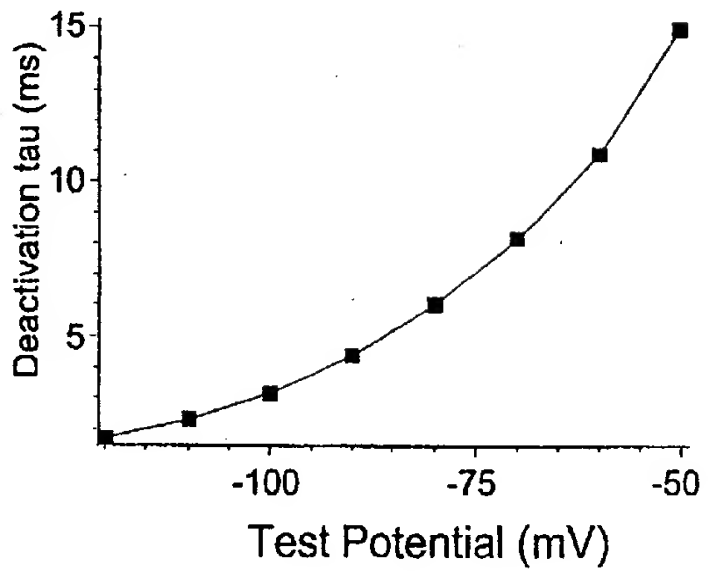

Fig. 8A

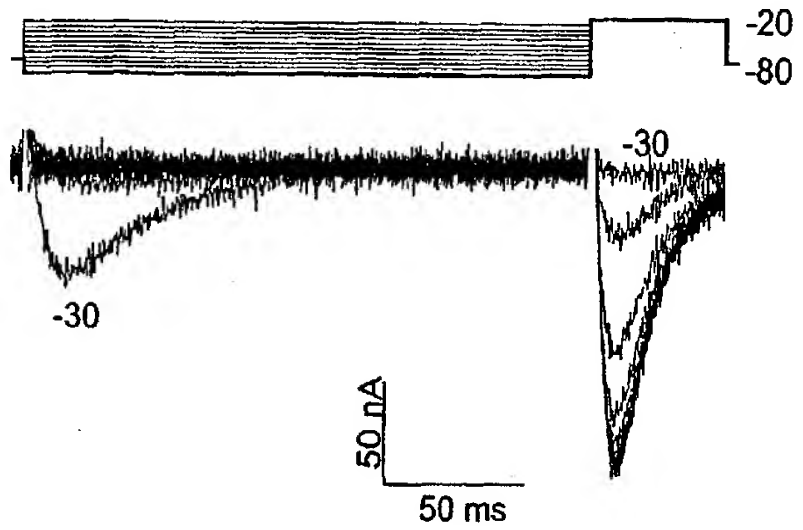

Fig. 8B

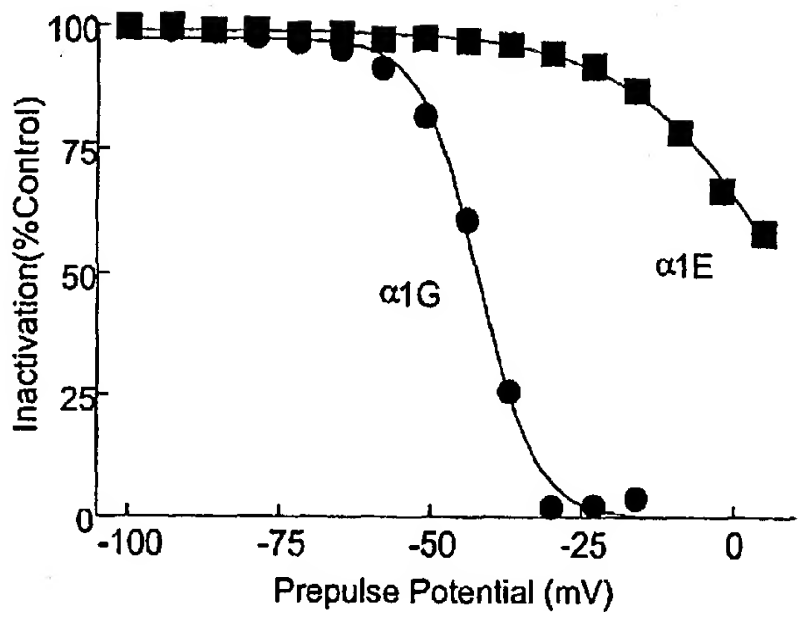

Fig. 8C

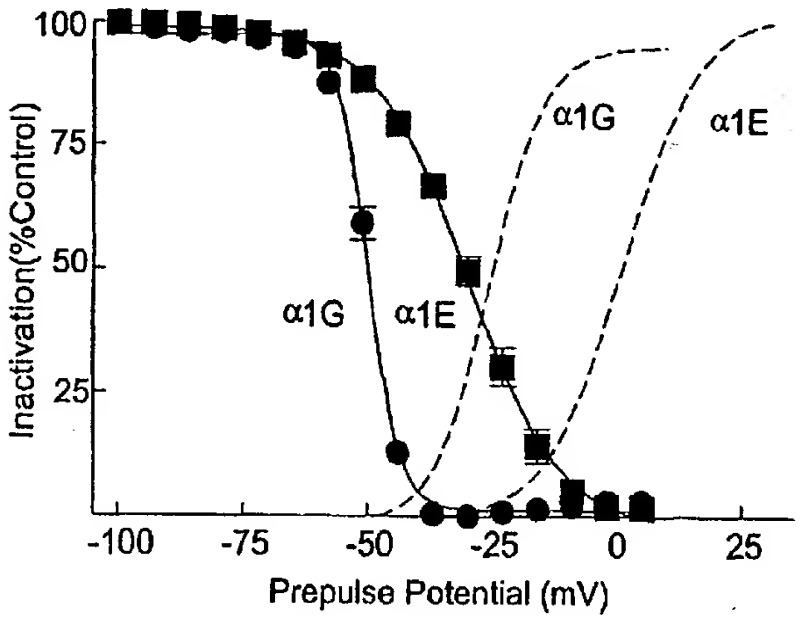

Fig. 9A

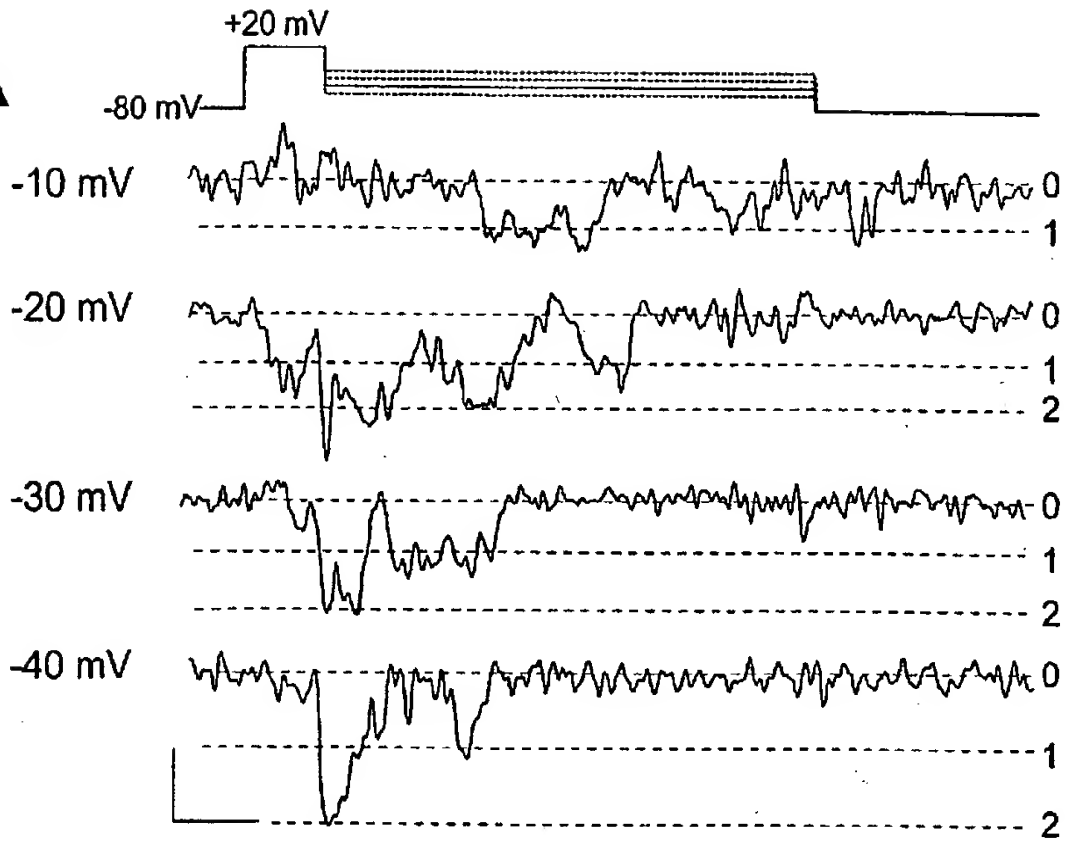

Fig. 9B

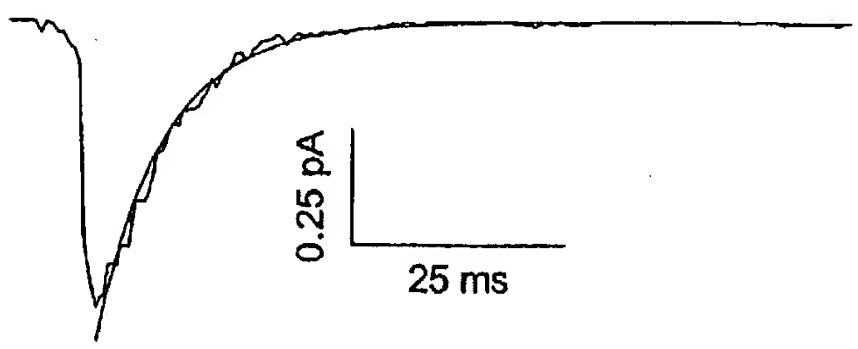

Fig. 9C

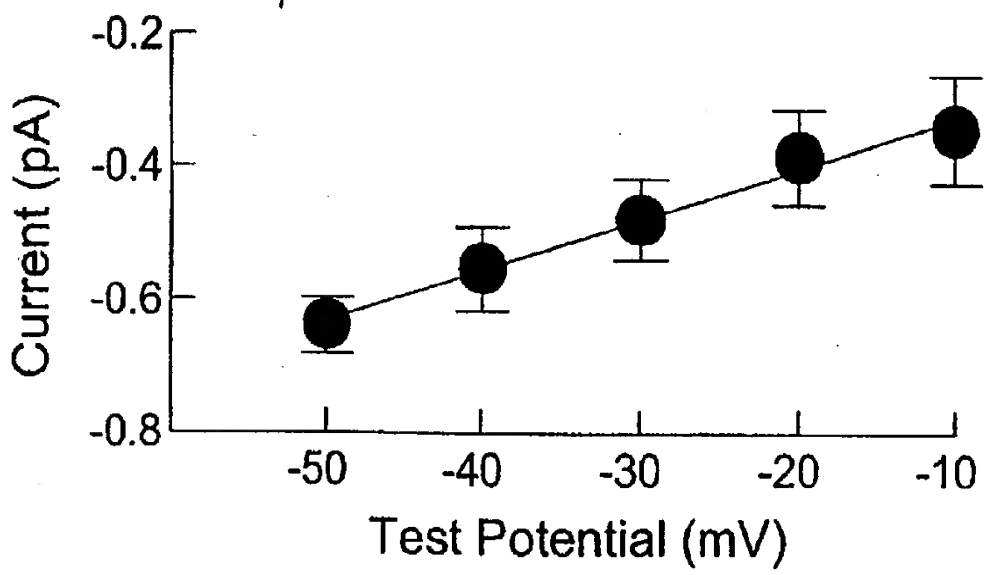

Fig. 10A

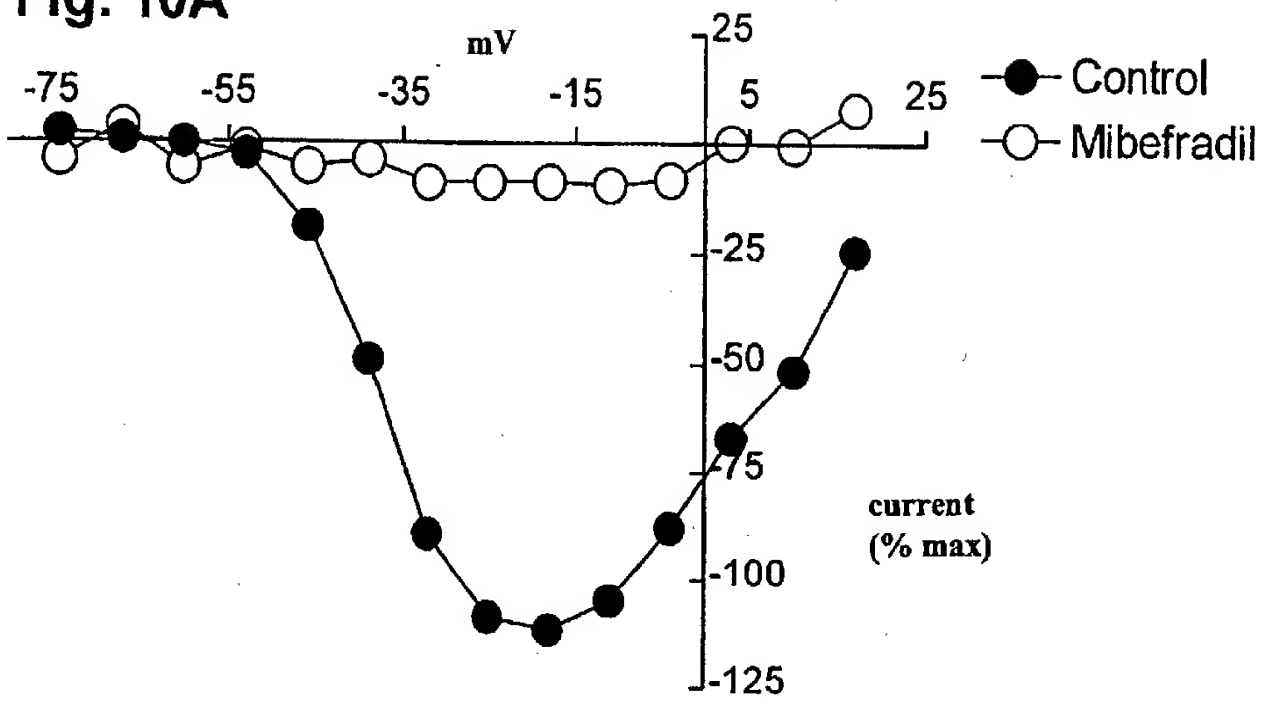

Fig. 10B